Kurdistan Journal of Applied Research (KJAR)
Print-ISSN: 2411-7684 | Electronic-ISSN: 2411-7706

Original Article

A Lightweight Distilled Model for Iris Recognition under a Subject-Disjoint, Periocular-Free Evaluation Protocol

Khalid Ibrahim Jajan [a, b *], **Wafaa Mustafa Abduallah** [c]

[a] Department of Information Technology, Akre University for Applied Science, Duhok, Iraq.
[b] Department of Information Technology, Duhok Polytechnic University, Duhok, Iraq.
[c] Department of Cybersecurity Engineering, Duhok Polytechnic University, Duhok, Iraq.

Submitted: 23 July 2026
Revised: 07 November 2026
Accepted: 15 November 2026

*** Corresponding Author**:
khalid.ibrahim@auas.edu.krd

Keywords: Iris recognition, Knowledge distillation, Occlusion-aware attention, Open-set biometric verification, Forensic biometrics.

How to cite this paper: K. I. Jajan, W. M. Abduallah, "A Lightweight Distilled Model for Iris Recognition under a Subject-Disjoint, Periocular-Free Evaluation Protocol", KJAR, vol. 11, no. 2, pp. 215-236, December 2026, doi: 10.24017/science.2026.2.12

Copyright: © 2026 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND 4.0) license

Abstract: Deep iris recognition on the Chinese Academy of Sciences Institute of Automation Iris-Thousand database is frequently evaluated on full eye images and without strict subject separation, thus, reported accuracy can be inflated both by periocular cues and identity leakage that overlap between partitions. This study aims to explore methods to addresses these confounds. The iris is isolated by learned semantic segmentation, rubber-sheet normalization and explicit eyelash and specular occlusion masking, and evaluation is restricted to a subject-disjoint, open-set five-fold protocol in which both eyes of a subject remain in the same partition. The recognition model is a compact EfficientNet-B0 with frequency-aware and occlusion-aware attention, generalized-mean pooling and a sub-center ArcFace objective, refined by knowledge distillation from a frozen ResNet-50 teacher. With seven-roll test-time augmentation (TTA), the distilled model attains an equal error rate (EER) of 1.242%, a verification rate of 98.49% at a false-acceptance rate of 1% and 95.25% rank-1 accuracy using 5.08 million parameters; without TTA the EER is 1.334%. The distilled model closes 60.4% of the student–teacher gap, attains a lower mean error than DenseNet-201 with about one quarter of its parameters, and is 2.4 times faster per image on the central processing unit than its teacher. Ablation and sensitivity analyses attribute the improvement to distillation rather than to the attention modules or to added capacity. With the iris and pupil erased, the periocular region alone attains an EER of 0.571% against 0.457% for the full eye frame, so full-frame evaluation on this database measures mainly non-iris appearance. In conclusion, iris-only, subject-disjoint evaluation is required for a valid comparison of iris matchers, and knowledge distillation offers an effective route to a compact and accurate model.

1. Introduction

The iris is among the most discriminative biometric traits available to forensic and civil identification. Its stochastic texture is fixed early in life, is largely independent between the same person's two eyes, and is protected behind the cornea; these properties have led to near-infrared iris matching being a foundation of national identity programs and criminal investigation alike. A recent systematic review of iris biometrics in forensic science confirms this role and identifies segmentation accuracy as a central limitation of current systems, since a defensible iris-only evaluation rests on this stage. Over the past decade, deep learning has substantially advanced iris feature extraction, replacing hand-crafted Gabor phase encoding with learned embeddings. However, two practical concerns have

received disproportionately little attention, and both bear directly on whether reported progress holds in actual use [1–3].

The first concerns the validity of evaluation. A large share of deep iris systems operates on the full eye image rather than on the segmented iris [4]. Segmentation accuracy is itself a limiting factor for recognition performance [5]. Full-eye inputs allow a convolutional network to exploit whatever is discriminative in the frame, including eyelids, eyelashes, skin texture, and the geometry of specular reflections. These periocular and acquisition-session cues are informative, but they are not iris texture [6], and a system that leans on them reports an error rate that understates its true iris-matching capability and degrades unpredictably under changes of sensor, illumination, or session. On Chinese academy of sciences institute of automation (CASIA) Iris-Thousand, a model restricted to the periocular region alone, with the iris masked out entirely, remains highly discriminative. Related practices compound the problem, as train/test partitions are frequently drawn over images rather than over subjects, permitting the same identity to appear on both sides of the split [1].

The second concern is efficiency. Forensic deployment increasingly targets resource-constrained edge hardware, where a matcher must run on a field device rather than a workstation [7, 8]. However, recent accuracy gains have come mainly from larger models. Model compression, including knowledge distillation, is an established way to reduce this cost in biometric systems [8]. In iris recognition, its use for identity embeddings remains limited.

This paper aims to addresses both concerns within a single framework. The iris is isolated using learned semantic segmentation followed by rubber-sheet normalization with explicit eyelash and specular occlusion masking, and evaluation is conducted exclusively under a subject-disjoint, open-set five-fold protocol with a quantitative leakage guard. On this foundation a compact EfficientNet-B0 is trained as an embedding network with frequency-aware and occlusion-aware attention, generalized-mean pooling, and a sub-center ArcFace objective. To combine the accuracy of heavy backbones with the small footprint required at the edge, a ResNet-50 teacher is then distilled into this lightweight student. Two distinct confounds are kept separate throughout this paper. Identity leakage is any overlap of subjects between the training and test partitions, and the subject-disjoint folds remove it. Periocular confounding is the use of non-iris cues surrounding the iris, and the segmentation and masking stage removes it. The two are reported separately because they have different causes and different remedies. The contributions of this study are the following:

- Periocular contamination in full-frame iris evaluation is quantified under a controlled protocol in which the input region is the only variable, and a segmentation-first pipeline free of that confound is derived from it.
- Knowledge distillation from a ResNet-50 teacher is shown to transfer a substantial part of the teacher's accuracy into a compact EfficientNet-B0 student without increasing its inference cost.
- A controlled analysis of the design space is provided, covering EfficientNet-B0, EfficientNet-B1, ResNet-50, DenseNet-201, and ConvNeXt-Tiny, an alternative teacher architecture, as well as component ablations and sensitivity analyses for the distillation weight and for the occlusion mask. Its negative outcomes are reported openly, since neither the attention modules nor added student capacity nor a more recent teacher architecture improved accuracy.
- All results are obtained under a single subject-disjoint, open-set protocol and are accompanied by four efficiency measures, which supports a controlled comparison that the existing literature does not provide.

The remainder of the paper is organized as follows. Section 2 reviews related works, section 3 details the materials and methods, including the experimental setup. Section 4 presents the results, including the ablation, sensitivity, efficiency and validity analyses, section 5 discusses the findings and their limitations; and section 6 presents the conclusions of the study and directions for future works.

2. Related Works

Deep learning has largely displaced hand-crafted iris encoding, yet several techniques now standard in face recognition, namely angular-margin losses, spectral and occlusion-aware attention, and knowledge distillation, have received limited attention in iris recognition. The studies reviewed in this section were identified by searching the Scopus, Clarivate Web of Science, and IEEE Xplore databases for articles published between 2020 and 2026, using the keywords "iris recognition" combined with "knowledge distillation," "lightweight network," "attention," "occlusion," "open-set verification," "CASIA-Iris-Thousand," and "forensic biometrics," together with the foundational works on which the proposed method is built.

Representative deep systems illustrate the breadth of this literature. Winston *et al.* [9] compared a basic convolutional neural network (CNN) with hybrid models in which the softmax output layer was replaced by k-nearest neighbors (KNN) or support vector machine (SVM) classifiers; the CNN–SVM hybrid achieved the highest accuracy (97.8%), whereas the CNN–KNN hybrid performed worse. Zambrano *et al.* [10] used the low-level layers of pre-trained image-classification CNNs as feature extractors without fine-tuning them on new iris datasets, so that recognition required only a single matching operation. Ahmad *et al.* [11] designed a deep CNN-based model that integrates face and facial-part detection with iris detection and tracking, targeting difficulties such as specular reflection, occlusion and glare from eyeglasses. None of these studies reported evaluation under a subject-disjoint, open-set protocol on iris-only input.

Nsaif *et al.* [12] showed that classical pipelines remain competitive on non-ideal imagery; they fused a multi-scale gray-level co-occurrence matrix with a multi-range circle Hough transform for pupil localization, followed by rubber-sheet normalization, log-Gabor encoding and Hamming-distance matching, reporting 87.68% and 94.47% identification accuracy on CASIA v4 Distance and the Multimedia University (MMU) V2 databases, respectively; however, that work reported equal error rates (EER) only graphically and as relative improvements over alternative segmentation methods, not as absolute values. The method was nonetheless hand-engineered and database-tuned and, more importantly, partitioned images rather than subjects into train and test, so a given identity could appear on both sides of the split. Across the recent literature, rubber-sheet mapping persists as the geometric front-end segmentation remains predominantly classical rather than learned, and explicit eyelash and specular occlusion handling is rarely integrated into the representation. Lightweight segmentation has been pursued in parallel with Sun *et al.* [13] proposing the lightweight segmentation network, a multitask model that jointly predicts the iris mask and the parameterized pupillary and limbic boundaries, using multiscale concatenation blocks and grouped spatial attention to reach competitive accuracy on three noisy iris datasets with only 2.2M parameters. Moreover, D'Alessandro *et al.* [14] applied a U-Net to heterogeneous multisource iris data. Such work is complementary to the present study, which treats the segmentation stage as fixed and quantifies its influence on recognition rather than optimizing it.

Li *et al.* [15] addressed efficiency, which is required for forensic edge deployment, through distillation. Their KD-Eye compressed a 31 M-parameter U-Net teacher into a 0.9 M-parameter student (34×) while preserving segmentation quality (intersection over union (IoU) of 0.896 against 0.912) at 160 frames per second on mobile hardware, which indicates that the accuracy of a large teacher can be transferred to an edge-deployable student. Its scope is nonetheless limited, because the student learned a dense segmentation map for virtual-reality eye tracking, not a discriminative identity embedding, and was evaluated on 30 self-collected subjects. For iris recognition specifically, Latif *et al.* [16] applied pruning together with knowledge distillation to reduce an iris recognition model, reporting a substantial reduction in model size at a small cost in accuracy.

Deng *et al.* [17] introduced the additive angular-margin objective ArcFace, which is now standard for discriminative embeddings in face recognition; its migration to the iris, however, remains incomplete. Luo *et al.* [18] proposed a closely related design that combined channel, spatial and co-attention over fused iris and periocular inputs with a softmax-plus-center objective and reported that attention improved the EER relative to an attention-free baseline. Its evaluation used CASIA-Iris-Distance, the University of Beira Interior Periocular (UBIPr) database and the Aleix and Robert face

database rather than CASIA-Iris-Thousand, so its error rates are not directly comparable with those of the present study. Two qualifications matter. First, the input fuses iris and periocular regions, so an unquantified share of the discriminative power derives from eyelids, lashes, skin, and reflection geometry, which inflates the apparent iris performance and prevents a direct comparison with iris-only systems. Second, it employed center loss rather than an angular margin. On the other hand, Gautam *et al.* [19] applied a supervised class-encoding autoencoder to iris presentation-attack detection with strong intra-database accuracy, confirming the maturity of deep iris analysis but addressing a different task with no recognition metrics. The combination of a sub-center angular-margin objective with spectral and occlusion-aware attention was not found in the iris recognition studies retrieved by this search. Nevertheless, the attention modules adopted here are retained as architectural components rather than claimed as a contribution of this study, because they did not improve accuracy in the present configuration.

Attention has also been applied to iris texture in forms other than the one adopted here. Wei *et al.* [20] represented each iris image as a probabilistic distribution rather than a deterministic feature vector, using uncertainty embedding and uncertainty-guided curriculum learning to reduce the influence of acquisition factors while Wei *et al.* [21] proposed contextual measures with two parallel branches that learn global and local correlations between iris regions, the global branch providing robustness to local occlusions. Wei *et al.* [22] also introduced a Gabor Trident Network that estimates device-specific bands to learn spectral-invariant features for matching near-infrared against visible-light images, while Jia *et al.* [23] proposed a convolutional network for iris recognition in less constrained environments, the key component of which is an interaction block that computes an affinity matrix among all pairs of high-level features, thereby learning second-order relationships that yield a more robust iris texture representation under severe noise. These studies indicate that the benefit of attention in iris recognition depends strongly on the form it takes.

Sun *et al.* [24] proposed IrisFormer, a dedicated vision-transformer framework for iris recognition that adopted relative position encoding for rotation invariance and randomly masked tokens during training to improve robustness to occlusion and reflection. Luo *et al.* [25] retained a ResNet-50 backbone on CASIA-Iris-Thousand and integrated a lightweight attention module that combines non-global interactive channel attention with lightweight spatial attention, trained under an optimized triplet loss; this design reduced the computation of the attention block relative to the convolutional block attention module (CBAM) and lowered the EER by up to 23.26% relative to the compared models. The authors identified distillation from a large teacher into a lightweight student as future work, which is the direction pursued in the present study. Two studies confirm the continuing interest in attention and in efficiency for iris recognition, and neither reported results under a subject-disjoint, open-set protocol [24, 25]. Most directly related to the present work, Huynh *et al.* [26] proposed MobiIris, a MobileNetV3-Large student trained with a dynamic-margin triplet loss, hard-negative mining and feature-level distillation from a ResNet-50 teacher, and evaluated on CASIA-Iris-Thousand under a subject-level split; they reported an EER of 1.409% with a model size of 12 megabytes. That study and the present one therefore addresses the same problem with the same teacher family, but differ in protocol, because MobiIris used a single 70:15:15 split and discarded approximately 6% of the images at preprocessing, whereas the evaluation reported here uses five subject-disjoint folds and reports the effect of excluded images explicitly.

Beyond iris-specific work, recent studies confirm the relevance of the same mechanisms across biometric modalities. Ramesh and Krishnaveni [27] combined attention-guided segmentation with a transformer encoder and Siamese metric learning under a subject-disjoint protocol, whereas, Lu *et al.* [28] applied multilevel attention-based knowledge distillation to multimodal biometric recognition; and Liu *et al.* [29] distilled a lightweight model for palmprint and finger-vein retrieval on resource-constrained devices. These works support the premise adopted here, namely that distillation is an effective route to compact biometric models, but none of them reported iris recognition under a subject-disjoint, open-set, multi-fold protocol. The segmentation front-end on which iris-only evaluation depends has likewise advanced: Liu *et al.* [30] extracted contour parameters jointly with the segmentation mask, and Muhammad *et al.* [31] guided localization and segmentation for unconstrained

acquisition. Li and Feng [32] localized the iris from a double-center formulation. These developments motivate treating segmentation quality as a variable in its own right. Huo *et al.* [33] pruned a U-Net++ architecture for the same task, which indicates that segmentation accuracy and computational cost are themselves design variables.

Protocol design is as consequential as architecture, and the reviewed evaluations are either subject-agnostic, admitting identity leakage [12], or confined to leave-one-subject-out testing on small self-collected corpora [15]. CASIA-Iris-Thousand, which is large and contains eyeglasses and specular reflections, was used by Luo *et al.* [25] and Huynh *et al.* [26], but neither reported subject-disjoint, multi-fold results. No recent study reports open-set, subject-disjoint, multi-fold evaluation on this benchmark with transparent EER, verification-rate and rank-1 reporting. Open-set operation has also been studied directly. Sun *et al.* [34] rejected unknown classes through an outlier-aware feature network, confirming that closed-set accuracy does not transfer to open-set conditions, and Sun *et al.* [35] fused multi-scale convolutional features with circular Gabor filters. Neither study, however, combined open-set operation with subject-disjoint partitioning and multi-fold reporting, which is the combination adopted here. Recent studies have also continued to refine the classical segmentation, normalization, and encoding pipeline. Rubio and Magnier [36] combined Canny edge detection and iris unwrapping with binarized statistical image features (BSIF) to improve the preprocessing of iris images. Sharma *et al.* [37] used histogram cut selection with recursive entropy discretization for segmentation and a genetic algorithm to optimize the classifier and Aliyu *et al.* [38] extracted features from the normalized iris with a two-dimensional Chebyshev wavelet filter and classified them with a convolutional neural network, reporting a recognition accuracy of 98.97% on the CASIA Iris Image Database Version 4.0 (CASIA-IrisV4), the second version of the University of Beira Interior iris database and MMU. None of these studies, however, evaluated a compact distilled model under a subject-disjoint, open-set protocol.

The present study closes these gaps within one framework, and is evaluated on iris-only input under a subject-disjoint open-set five-fold protocol, which is stricter than that of Luo *et al.* [18], whose results benefit from periocular cues. Therefore, the reported operating point is a conservative measurement free of identity leakage and periocular confounding rather than one inflated by non-iris signal. Table 1 summarizes the reviewed studies and contrasts their approach, input region, use of knowledge distillation, and evaluation protocol with those of the present work.

Table 1: Summary of related iris recognition studies.

Study	Year	Approach	Input	KD	Evaluation protocol	Limitation addressed here
[9]	2022	Basic CNN and hybrid CNN with KNN/SVM classifiers	Cropped eye image (periocular included)	No	Closed-set; augmented images split into train/test (IIT Delhi)	Closed-set; no open-set verification
[10]	2022	Early-layer filters of pre-trained CNNs without training; single matching	Iris (normalized)	No	NR	No learned embedding; no distillation
[11]	2024	integrated deep-learning iris detection and tracking	NR	NR	NR	Detection and tracking, not recognition
[12]	2022	Hand-crafted: GLCM + Hough; log-Gabor; Hamming	Full-eye	No	Image-level split	Not subject-disjoint
[13]	2022	Multitask lightweight segmentation network (2.2 M)	Eye image	No	NR	Segmentation, not recognition
[14]	2023	Modified U-Net	Eye image (heterogeneous)	No	NR	Segmentation, not recognition
[15]	2024	U-Net teacher distilled to 0.9 M student	Near-eye	Yes	Leave-one-subject-out	Segmentation, not recognition
[16]	2024	Pruning + knowledge distillation	NR	Yes	NR	Evaluation protocol not reported
[17]	2022	Additive angular-margin loss (ArcFace)	Face	No	Face benchmarks	Face, not iris
[18]	2021	CNN; channel, spatial and co-attention; center loss	Iris + periocular	No	Not subject-disjoint	Periocular cues admitted

Table 1: Continued.

[19]	2022	FCN-DAE + CNN	Iris (cropped)	No	2-fold, repeated	Presentation-attack detection
[20]	2022	Uncertainty embedding; uncertainty-guided curriculum learning	NR	No	NR	No distillation
[21]	2022	Contextual measures (global and local branches)	NR	No	Within- and cross-database	No distillation
[22]	2022	Gabor Trident Network for spectral-invariant features	NIR and VIS iris	No	Cross-spectral	Cross-spectral task; no distillation
[23]	2022	Second-order ConvNet; ordinal encoding; mask network	NR	No	NR	No distillation
[24]	2025	Vision transformer; token masking	Iris	No	Not subject-disjoint	No open-set protocol
[25]	2025	ResNet-50; lightweight attention; triplet	Iris	No	Image-level 4:1 train/test split; subject-disjointness not stated	Distillation left as future work
[26]	2026	MobileNetV3-L distilled from ResNet-50	Iris	Yes	Single 70:15:15 subject-level split	Single split; 6% of images discarded
[27]	2026	Denoising autoencoder + attention U-Net + ViT Siamese	Iris (segmented)	No	Subject-disjoint verification	No distillation
[28]	2025	Multilevel parallel attention distillation	Fingerprint + face	Yes	NR	Not iris
[29]	2025	Deep distillation hashing	Palmprint + finger vein	Yes	Retrieval	Not iris
[30]	2023	Multitask segmentation with Swin-Transformer attention; contour parameters	NR	No	NR	Segmentation, not recognition
[31]	2023	Guided localization and segmentation network	Eye image (unconstrained)	No	NR	Segmentation, not recognition
[32]	2023	Double-center localization + mask segmentation	Eye image (VIS)	No	Localization/ segmentation IoU (NICE-II, MICHE)	Segmentation, not recognition
[33]	2022	Attention U-Net++ with EfficientNetV2 blocks; pruning	Eye image (NIR and VIS)	No	NR	Segmentation, not recognition
[34]	2022	Outlier-aware open-set network	Iris	No	Open-set, single split	Not subject-disjoint; no folds
[35]	2022	Local circular Gabor augmentation + multi-scale CNN feature fusion	Iris (normalized)	No	Closed-set softmax classification (CASIA-Iris-Syn, CASIA-Iris-Lamp)	No distillation
[36]	2024	Canny edges + iris unwrapping; BSIF features	Iris (normalized)	No	Not stated as subject-disjoint	No distillation; no open-set protocol
[37]	2025	Histogram cut selection; entropy discretization; GA-optimized classifier	Iris	No	Recognition accuracy; not stated as subject-disjoint	No distillation; no open-set protocol
[38]	2026	2D Chebyshev wavelet filter + CNN classifier	Iris (normalized)	No	Recognition accuracy (closed-set)	No distillation; no open-set protocol
This work	—	**EfficientNet-B0 distilled from ResNet-50**	**Iris only**	**Yes**	**Open-set, subject-disjoint, five-fold**	—

BSIF = binarized statistical image features; CNN = convolutional neural network; ConvNet = convolutional network; FCN-DAE = fully convolutional denoising autoencoder; GA = genetic algorithm; GLCM = gray-level co-occurrence matrix; IoU = intersection over union; KD = knowledge distillation; KNN = k-nearest neighbors; M = million; NIR = near-infrared; NR = not reported; SVM = support vector machine; ViT = Vision Transformer; VIS = visible light; 2D = two-dimensional.

3. Materials and Methods

Building on the gaps identified above, the proposed system is based on the premise that reliable iris recognition requires the iris to be isolated before it is represented, and that the accuracy of a large network can be transferred into a compact one without inheriting its cost. Each near-infrared acquisition is treated as a normalized iris sample $u \in \mathbb{R}^{H\times W}$ with an occlusion mask $o \in \{0,1\}^{H\times W}$ and an

identity label $y \in \{1, \ldots, C\}$, where the left and right eyes are counted as separate classes. The database contains 2,000 such eye-classes; because every fold train on four fifths of the subjects, the prototype layer of a single fold has $C = 1{,}600$ classes. The goal is an embedding F_θ: $(u, o) \mapsto e \in \mathbb{R}^d$ under which, using cosine similarity, genuine comparisons score higher than impostor comparisons on identities never seen in training.

Stage 1 isolates the iris, whereby the near-infrared image is segmented by a multilabel U-Net++/MobileNetV2 network; the pupillary and limbic boundaries are fitted as equivalent-area circles; the annulus is unwrapped by rubber-sheet normalization into a 128×512 strip; and eyelash and specular pixels are masked and retained as an occlusion mask. Stage 2 encodes the strip with a lightweight EfficientNet-B0 augmented by frequency-aware and occlusion-aware attention and generalized-mean pooling, yielding a 512-D, ℓ_2-normalized embedding. Stage 3 is the training objective: the student is supervised by a sub-center ArcFace loss and, simultaneously, by a cosine knowledge-distillation (KD) loss that aligns its embedding with that of a frozen ResNet-50 teacher consuming the same normalized strip, and gradients update the student only. Stage 4 evaluates the trained model under a subject-disjoint, open-set five-fold protocol with TTA and a leakage guard. Figure 1 summarizes the four stages of the pipeline.

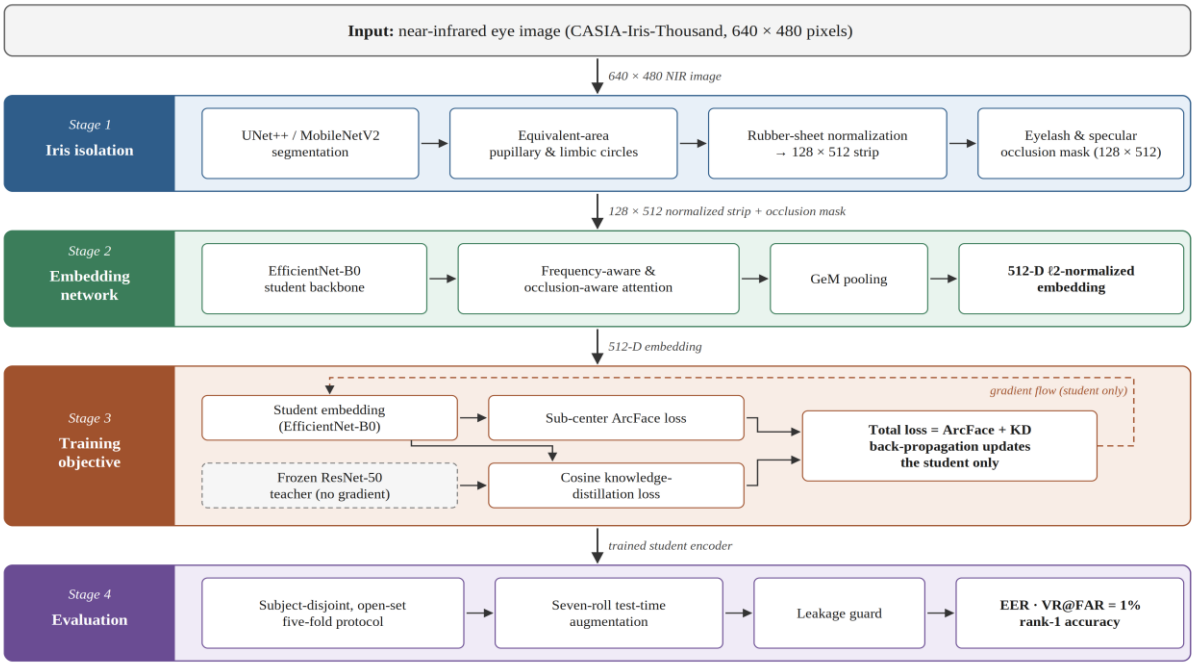


Figure 1: Overview of the proposed pipeline, from iris isolation to open-set evaluation. CASIA = Chinese Academy of Sciences Institute of Automation; D = dimensional; EER = equal error rate; FAR = false-acceptance rate; GeM = generalized-mean pooling; VR = verification rate.

3.1. Segmentation and Rubber-Sheet Normalization

A multilabel semantic segmenter assigns every pixel independent membership to the eyeball, iris, pupil and eyelash classes. Pupillary and limbic boundaries are recovered as equivalent-area circles of radius $r = \sqrt{(A/\pi)}$, where A is the area of the corresponding region; this formulation is robust to the small specular holes typical of the database. The iris annulus is then mapped to a rectangular strip of fixed size using Daugman's rubber-sheet model [39]. Each point of the strip, with radial coordinate $\varrho \in [0, 1]$ and angular coordinate $\theta \in [0, 2\pi)$, takes its value from the image point that lies at the fraction ϱ of the distance between the pupillary boundary and the limbic boundary along the direction θ, as given in equation (1)

$$x(\rho, \theta) = (1 - \rho)\, x_p(\theta) + \rho\, x_l(\theta), \qquad y(\rho, \theta) = (1 - \rho)\, y_p(\theta) + \rho\, y_l(\theta) \tag{1}$$

where (x_p, y_p) **and** (x_l, y_l) trace the pupillary and limbic circles, ϱ the normalized radial coordinate and θ the angular coordinate. Pixels labeled as eyelash, non-iris, or specular reflection are collected into **o** and set to zero, so that only genuine iris texture propagates downstream. Figure 2 illustrates this stage.

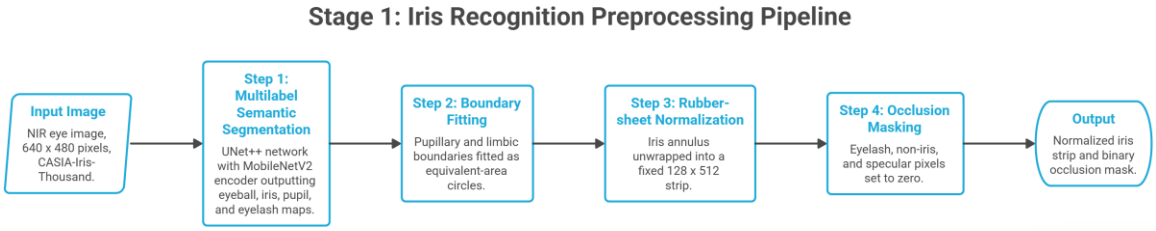


Figure 2: Iris isolation stage of the pipeline.

3.2. *Feature Extraction*

A lightweight backbone yields a feature tensor $\mathbf{F} \in \mathbb{R}^{\mathbf{C}\prime\times\mathbf{h}\times\mathbf{w}}$, refined by two complementary attention mechanisms. Because iris texture is intrinsically a frequency phenomenon, a frequency-aware module rescales each channel by its spectral energy as given in equation (2)

$$a = \sigma\left(\mathrm{MLP}\left(1/hw \sum_{u,v} |\mathcal{F}(F)|\right)\right) \qquad F' = F \odot a \qquad (2)$$

where $\mathcal{F}(\cdot)$ is the two-dimensional discrete Fourier transform, F the backbone feature tensor with C′ channels and spatial size h × w, (u, v) the spatial position within the feature map, MLP a two-layer perceptron, the vector of channel attention weights, F′ the frequency-refined tensor, $\odot$ the element-wise product and $\boldsymbol{\sigma}$ the sigmoid. An occlusion-aware module then suppresses unreliable regions by conditioning a spatial gate, as given in equation (3), on the pooled valid map $\bar{\mathbf{o}} = \mathbf{1} - \mathbf{pool}(\mathbf{o})$

$$F'' = F' \odot \sigma\left(\phi\left(\left[F'_{avg};\ F'_{max};\ \bar{o}\right]\right)\right) \qquad (3)$$

where pool(·) resizes the occlusion mask o to the h × w resolution of the feature map by pooling, ō is the resulting map of valid pixels, F″ is the attention-refined tensor, the square brackets [;] denote concatenation along the channel dimension, F'_{avg} and F'_{max} are the average and the maximum of F′ taken across channels at each spatial position, φ is a convolutional spatial gate, following the spatial attention of CBAM [40], and $\odot$ is the element-wise product; masked pixels therefore contribute little to the descriptor. Generalized-mean pooling [41] aggregates the refined map, and a linear projection with batch normalization yields the embedding e, as given in equation (4)

$$z = (1/hw \sum_i (F''_i)^p)^{1/p} \qquad e = \mathrm{BN}(Wz) \qquad \hat{e} = e/\| e \|_2 \qquad (4)$$

where z is the pooled descriptor, i indexes the h × w spatial positions, p the learnable generalized-mean exponent, ê the ℓ_2-normalized embedding, W a linear projection, and BN denotes batch normalization. Figure 3 shows the complete embedding network.

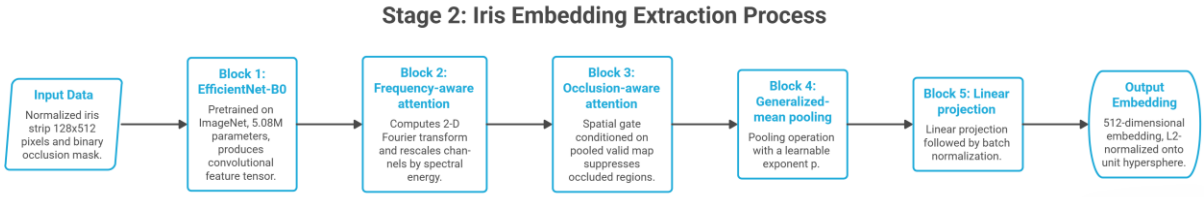


Figure 3: Embedding network stage of the pipeline.

3.3. *Training Objective with Angular Margin and Knowledge Distillation*

Each class is represented by K unit-norm prototypes, and class similarity is the strongest sub-center response, cos θ_j = max_k ($\hat{e}^T \hat{W}_{j,k}$), where ê is the ℓ_2-normalized embedding, Jok is the k-th normalized prototype of class j, and the maximum is taken over the K sub-centers. Following the sub-

center formulation [42] that extends the additive angular-margin softmax [17], the loss enforces an additive margin m on the target class under scale s, as given in equation (5)

$$\mathcal{L}_{\text{arc}} = -\log \frac{\exp\left(s\cos\left(\theta_y + m\right)\right)}{\exp\left(s\cos\left(\theta_y + m\right)\right) + \sum_{j \neq y} \exp\left(s\cos\theta_j\right)} \qquad (5)$$

where s is the scale, m the additive angular margin, θ_y the angle between the embedding and the prototype of the true class, θj the angle to the prototype of any class j, and C the number of classes of the training fold, which is 1,600.

Allowing several sub-centers per identity lets atypical samples, such as images with eyeglasses or changing reflections, attach to their own prototype rather than distorting a single class center.

Although ArcFace was introduced for face recognition [17], the objective is independent of the backbone that produces the embedding, because it operates only on the $\ell 2$-normalized embedding and a set of normalized class prototypes. Any network that outputs a d-dimensional embedding can therefore be trained with it unchanged, which is why the identical objective is applied to all five backbones compared in this study. Its suitability for iris follows from the same geometric principle that motivates it for faces, namely the maximization of inter-class angular separation on the unit hypersphere, while the sub-center extension absorbs iris-specific intra-class variation such as eyeglasses and specular reflections.

The second component of the training objective entails a high-capacity teacher $\mathbf{T}$, trained beforehand under exactly the same segmentation, normalization and evaluation protocol, who is frozen and supervises the compact student at the level of the embedding. Given the same input, teacher and student produce the normalized descriptors given in equation (6)

$$\hat{e}_s = \frac{F_\theta(u,o)}{\|F_\theta(u,o)\|_2}, \qquad \hat{e}_t = \frac{T(u,o)}{\|T(u,o)\|_2} \qquad (6)$$

Where, F_θ is the student and T the frozen teacher. When the two embeddings differ in dimension, a learnable projector, $g\phi$, maps the student embedding to the dimension of the teacher embedding, and the projected vector is again $\ell 2$-normalized. Alignment is imposed in the angular domain, as given in equation (7)

$$\mathcal{L}_{\text{kd}} = 1 - \hat{e}_t^{\top} \tilde{e}_s = 1/2 \parallel \tilde{e}_s - \hat{e}_t \parallel_2^2 \qquad (7)$$

The second equality in equation (7), which holds exactly for unit-norm vectors, and where $\tilde{\mathbf{e}}_s$ **is projected**, **normalized student embedding**, shows that matching cosine direction is equivalent to minimizing squared Euclidean distance on the unit hypersphere, and therefore optimizes precisely the geometry the cosine score exploits at test time. The student is trained on the combined loss as given in equation (8)

$$\boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{L}}_{\mathbf{arc}} + \boldsymbol{\lambda}\, \boldsymbol{\mathcal{L}}_{\mathbf{kd}} \qquad (8)$$

where λ weights the distillation term against the angular-margin term, and θ and ϕ denote the parameters of the student network and of the projector, respectively.
The total loss is minimized only with respect to $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ while the teacher receives no gradient, $\boldsymbol{\nabla}_{\mathbf{T}}\boldsymbol{\mathcal{L}} = \mathbf{0}$. The hard-label term preserves class separability; the distillation term injects the richer inter-sample structure that a small network cannot discover unaided.

Training proceeds as follows. The teacher is first trained under the same pipeline with the sub-center ArcFace loss alone and then frozen. For each mini-batch, the normalized strips and occlusion masks are passed through both networks; the student's embedding enters the sub-center ArcFace loss with the ground-truth identities and, after projection, the cosine distillation loss against the teacher's embedding of the same input. The two terms are combined as in equation (8) and back-propagated through the student and the projector only, so the teacher is never updated. The weight λ is set to 1, the standard equal weighting of the two terms, which is justified a priori because both terms operate in the same angular space at comparable scales. Figure 4 summarizes the training objective and the gradient flow.

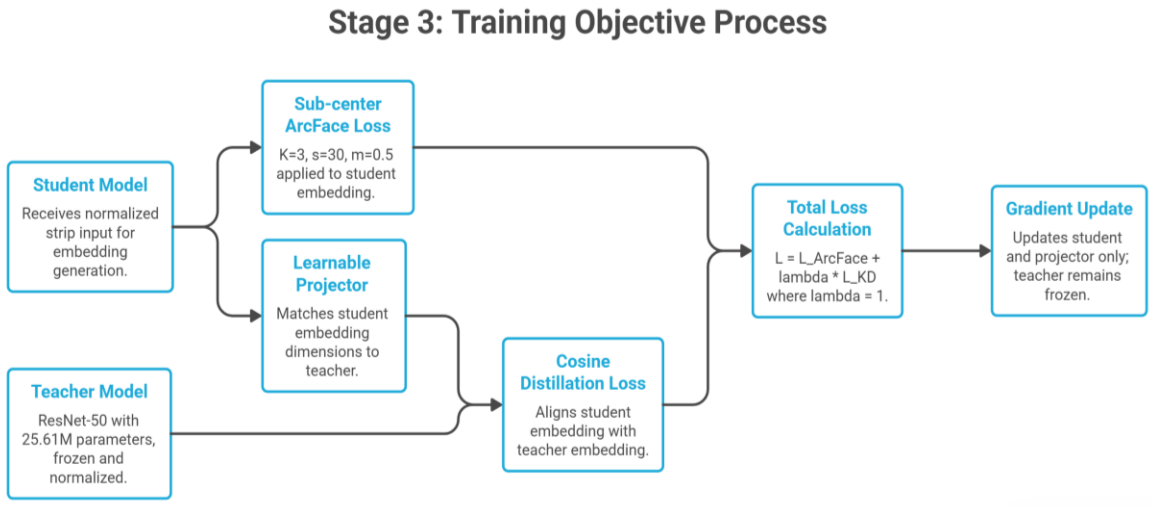


Figure 4: Training objective stage of the pipeline.

3.4. Open-Set Verification Protocol

Identities are divided into N (= five) subject-disjoint folds so that partitioning is performed over the 1,000 subjects rather than over the 2,000 eye-classes. Both eyes of a subject, and all their images, therefore fall entirely within either the training or the test partition, which guarantees that verification is measured on unseen identities and precludes the periocular- and session-based leakage that inflates full-frame protocols. A comparison between samples a and b is scored by the cosine similarity sab = $\hat{e}a^{\top}\hat{e}b$, where $\hat{e}a$ and $\hat{e}b$ are the ℓ_2-normalized embeddings of the two samples; the embeddings are averaged over R angular-roll augmentations and renormalized before scoring. Performance is summarized by the EER at the threshold τ^* at which the false-acceptance rate (FAR) and false-rejection rate (FRR) coincide, EER = FAR(τ^*) = FRR(τ^*), complemented by the verification rate at FAR = 1% and rank-1 accuracy. As a safeguard, the genuine–impostor separation is measured by the decidability index [39], as given in equation (9)

$$d' = \frac{\mu_{gen} - \mu_{imp}}{\sqrt{1/2\left(\sigma_{gen}^2 + \sigma_{imp}^2\right)}} \qquad (9)$$

where $\boldsymbol{\mu_{gen}}$, $\boldsymbol{\sigma_{gen}^2}\sigma_{gen}$ and $\boldsymbol{\mu_{imp}}p$, $\boldsymbol{\sigma_{imp}^2}$ are the means and standard deviations of the genuine and impostor score distributions. Any operating point yielding EER < 0.5% or d′ > 8 is flagged as implausibly strong and audited for leakage before acceptance.

Each test partition therefore holds 200 unseen subjects: 400 unseen eye-classes with each training partition holding 800 subjects, which is 1,600 eye-classes. Because the two eyes of one subject share periocular, illumination and eyeglass characteristics, assigning them to the same partition is required for the open-set claim.

The number of folds is set to five. With 2,000 eye-classes, this places 400 unseen eye-classes in each test partition and 1,600 in each training partition, which provides enough test identities for a stable EER estimate and enough training identities to learn discriminative embeddings. Fewer folds would shrink the training pool, whereas ten-fold or leave-one-subject-out evaluation would increase computation substantially with negligible variance reduction at this scale. Figure 5 summarizes the evaluation protocol.

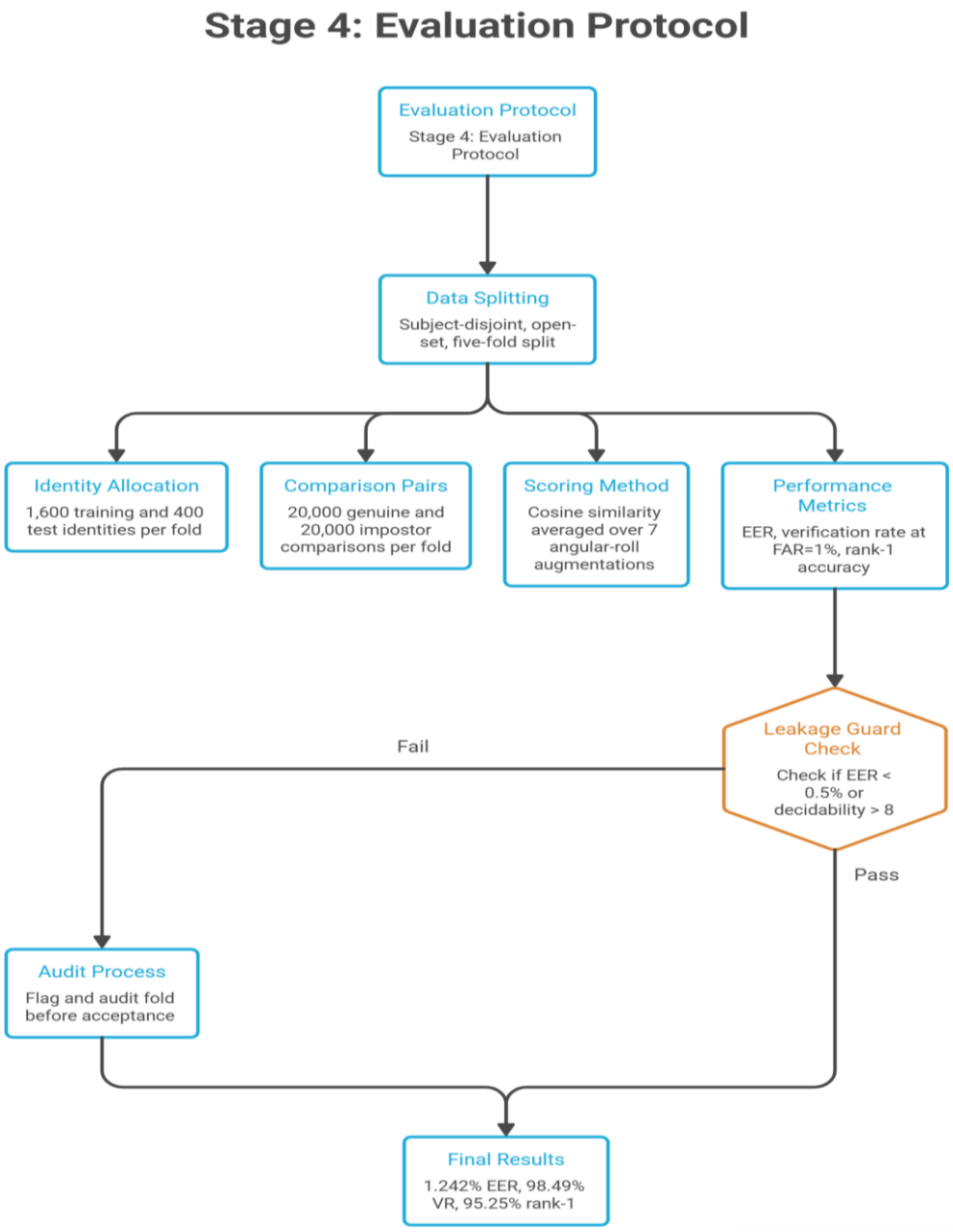


Figure 5: Evaluation protocol stage of the pipeline.

3.5. Experimental Setup

All experiments use the CASIA-Iris-Thousand subset of the CASIA-IrisV4 [43], which is publicly available for research after registration at http://biometrics.idealtest.org/. The subset contains 20,000 near-infrared images of 640×480 pixels from 1,000 subjects, acquired with an IKEMB-100 dual-eye iris camera, with 10 images of the left eye and 10 images of the right eye of each subject. Because a person's left and right irises have statistically largely independent patterns, each eye is treated as a distinct identity, which gives 2,000 eye-classes of 10 images each. The main sources of intra-class variation in CASIA-Iris-Thousand are eyeglasses, specular reflections, and partial eyelid occlusion. Table 2 summarizes the statistics of the database and of the partitions used in this study.

Table 2: Statistics of the CASIA-Iris-Thousand database used in this study.

Property	Value
Source database	CASIA-IrisV4, subset CASIA-Iris-Thousand [43]
Provider	Chinese Academy of Sciences Institute of Automation (CASIA)
Acquisition device	IKEMB-100 dual-eye iris camera (IrisKing)
Illumination	Near-infrared
Image resolution	640 × 480 pixels
Subjects	1,000
Eye-classes (identities)	2,000 (left and right eye of each subject)
Images per eye-class	10
Total images	20,000
Images segmented successfully	19,954 (99.8%)
Eye-classes per fold (training / test)	1,600 / 400
Main intra-class variations	Eyeglasses, specular reflections, eyelid occlusion
Access	Publicly available for research after registration (http://biometrics.idealtest.org/)

Iris isolation is performed once, offline, and cached, so every model consumes an identical representation. Images are segmented by a pretrained multilabel network (U-Net++ [44] with a MobileNetV2 [45] encoder), whose four channels are thresholded independently rather than by arg-max. Acquisitions whose limbic radius does not exceed the pupillary radius by a factor of 1.3 are rejected as degenerate. The annulus is unwrapped into a 128×512 strip; eyelash, non-iris and near-saturated specular pixels are zeroed and retained as an occlusion mask. Segmentation succeeds on 19,954 of 20,000 images (99.8%), and all 2,000 identities retain at least two usable samples, so no identity is lost from the verification protocol. Strips are resized to 256×256 at the network input. The 46 unsegmentable images (0.23%), mainly acquisitions with severe eyeglass reflections, are treated as failure-to-acquire cases, which in deployment would be rejected and re-acquired rather than matched.

The proposed student is an ImageNet-pretrained EfficientNet-B0 [46] with 5.08M inference parameters (backbone and embedding head, excluding the training-only classification layer). For a controlled study of the accuracy–capacity relationship, EfficientNet-B1 (7.58 M), DenseNet-201 (20.00M) and ResNet-50 (25.61M) are trained under an otherwise identical pipeline; ResNet-50 additionally serves as the distillation teacher. ConvNeXt-Tiny [47] (approximately 28.6M) is trained as an alternative teacher to test whether a more recent architecture yields a stronger teacher. All backbones share the same head, so performance differences are attributable to the backbone alone.

Models are optimized with AdamW [48], which decouples weight decay from the adaptive gradient update, using an initial learning rate of 3×10^{-4} and a weight decay of 10^{-4}. The learning rate follows a cosine annealing schedule [49] over 40 epochs, so that larger steps are taken early in training and smaller steps near convergence. The same learning rate is applied unchanged to every backbone rather than tuned for each architecture, so that no backbone benefits from a configuration specific to it. Training uses a batch size of 64 and automatic mixed precision on a single NVIDIA Tesla T4 graphics processing unit (GPU). The objective is sub-center ArcFace with three sub-centers per class, s=30, m=0.5. Augmentation comprises random circular shifts of the strip, emulating iris rotation, and mild brightness jitter. For distillation, the fold-matched ResNet-50 teacher is frozen and λ=1.

Identities are partitioned into five subject-disjoint folds. Each test fold samples 20,000 genuine and 20,000 impostor pairs under a fixed seed. The EER, the verification rate (VR) at a FAR of 1% (VR@FAR=1%) and the rank-1 accuracy are given as the mean ± standard deviation over the five folds. Test-time augmentation) averages the embeddings of seven angular rolls. All backbones and variants are evaluated on identical folds, pairs and seeds, so all comparisons are made under identical conditions. Efficiency is quantified by four measures computed identically for every model, namely the number of inference parameters; the computational cost in giga floating-point operations (GFLOPs) for one 256×256 input, excluding the training-only classification head; the stored model size in megabytes

(MB) at 32-bit floating-point precision (FP32); and the inference latency for a single image on the Tesla T4 GPU and on the host central processing unit (CPU) averaged over repeated runs after warm-up.

4. Results

This section reports the results obtained under the subject-disjoint, open-set five-fold protocol. All recognition configurations, except the full-frame validity analysis of the input region, were generated end-to-end with the single pipeline described above, under fixed random seeds, folds and verification pairs. The results are presented in seven parts, namely the verification performance of the compared backbones, an ablation of the design choices, the sensitivity to the distillation weight and to the teacher architecture, the computational efficiency, the robustness to mask quality and segmentation failures, a validity analysis of the input region, and the stability of the results across folds. Unless stated otherwise, all values are five-fold means ± standard deviations.

4.1. Verification Performance of the Compared Backbones

Table 3 reports the verification performance of the five models and all rows share the same segmentation, normalization, embedding head, folds, verification pairs and random seeds; the results are five-fold means with standard deviations, and the verification rate and rank-1 accuracy are reported with TTA.

Among the models trained without distillation, the two largest networks attain the lowest errors. ResNet-50 attains the lowest error of any single model (1.145% ± 0.088%), ahead of DenseNet-201 (1.303% ± 0.135%) and the compact EfficientNet-B0 (1.390% ± 0.067%). Under this protocol, the lightweight backbone therefore does not outperform the larger networks.

Increasing the capacity of the student does not reduce its error. EfficientNet-B1 (7.58 M) reaches 1.560% ± 0.116%, a higher error than the smaller EfficientNet-B0 (5.08M), and this ordering was consistent across the independent runs performed. All backbones share identical hyperparameters, and the learning rate and regularization were not tuned separately for each architecture. The result therefore shows only that added capacity did not reduce error under a common training budget, and it does not establish a causal relation.

Knowledge distillation from the frozen ResNet-50 teacher lowers the EER of the EfficientNet-B0 student (5.08M parameters) from 1.390% ± 0.067% without distillation to 1.242% ± 0.074%. This reduction of 0.148 percentage points closes 60.4% of the 0.245-point gap between the baseline student (1.390%) and the teacher (1.145%). The distilled student has a lower mean error than DenseNet-201 (1.303%) with approximately one quarter of its parameters, and its EER is 0.097 percentage points higher than that of the teacher, which has five times as many parameters. Figure 6 shows this accuracy–efficiency frontier and figure 7 isolates the gap-closing effect.

Table 3: Verification performance of the compared backbones.

Model	Role	Params (M)	EER (%)	EER + TTA (%)	VR@FAR=1% (%)	Rank-1 (%)
ResNet-50	Teacher	25.61	1.248 ± 0.113	1.145 ± 0.088	98.71	95.90
DenseNet-201	Baseline	20.00	1.416 ± 0.173	1.303 ± 0.135	98.45	95.17
EfficientNet-B1	Capacity probe	7.58	1.700 ± 0.133	1.560 ± 0.116	97.83	93.53
EfficientNet-B0	Student (baseline)	5.08	1.475 ± 0.076	1.390 ± 0.067	98.20	94.52
EfficientNet-B0 + KD	**Proposed**	**5.08**	**1.334 ± 0.097**	**1.242 ± 0.074**	**98.49**	**95.25**

EER = equal error rate; FAR = false-acceptance rate; KD = knowledge distillation; M = million; TTA = test-time augmentation; VR = verification rate.

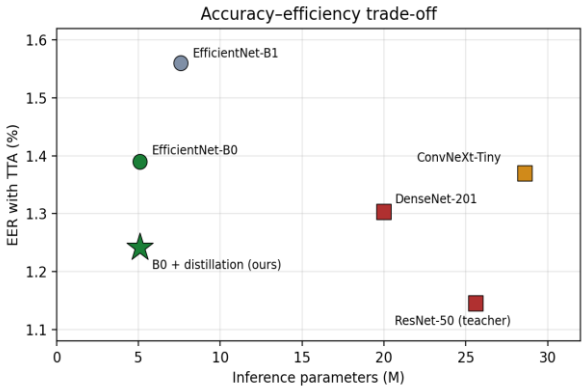


Figure 6: Equal error rate against the number of inference parameters.

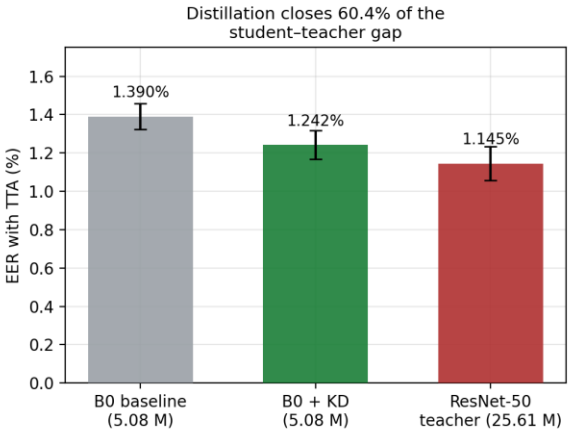


Figure 7: Equal error rate of the baseline student, the distilled student and the teacher.

4.2. Ablation Study

Table 4 isolates each architectural component, namely frequency-aware attention (FAA) and occlusion-aware attention (OAA), and the distillation step. All configurations use the identical pipeline, folds and seeds. The change in EER given in table 4 is computed with TTA relative to the full model.

The two attention modules contribute little on their own. Removing frequency-aware attention raises the EER by only 0.018 percentage points, well within the cross-fold standard deviation, and removing occlusion-aware attention lowers it slightly (−0.023 percentage points). Removing both modules, however, raises the EER by 0.064 percentage points, more than the sum of the individual effects, indicating that the two modules are partly redundant, as each compensates in part for the absence of the other. The occlusion-aware result is reported as a negative finding. A likely reason is that the segmentation stage already sets eyelash and specular pixels to zero before the embedding network, which leaves the occlusion-aware module little additional information to use.

The capacity probe (EfficientNet-B1) is likewise negative (+0.170 percentage points). Distillation, by contrast, reduces the error by 0.148 percentage points at an unchanged parameter count, which is the largest effect in the table and more than twice the combined contribution of both attention modules.

Table 4: Ablation of the design choices.

#	Configuration	EER + TTA (%)	Δ EER (pp)
1	Full model (FAA + OAA)	1.390 ± 0.067	—
2	w/o frequency-aware attention	1.408 ± 0.107	+0.018
3	w/o occlusion-aware attention	1.367 ± 0.021	−0.023
4	w/o both attention modules	1.454 ± 0.125	+0.064
5	Increased capacity (EfficientNet-B1)	1.560 ± 0.116	+0.170
6	**+ knowledge distillation (ResNet-50 → B0)**	**1.242 ± 0.074**	**−0.148**

EER = equal error rate; FAA = frequency-aware attention; OAA = occlusion-aware attention; pp = percentage points; TTA = test-time augmentation.

4.3. Sensitivity to the Distillation Weight and the Teacher Architecture

Table 5 reports the sensitivity of the distilled student to the weight λ in equation (8), with the teacher, folds and seeds held fixed. λ=2.0 gives a lower mean error (1.185%) than λ=1.0 (1.242%), but with a 2.5-fold higher cross-fold standard deviation (0.185 versus 0.074 percentage points), and the two ranges overlap; λ=0.5 is worse on both counts. The value λ=1, which was fixed before the experiments, is therefore kept because it gives the most stable performance across subject partitions, and the full sweep is given so that the alternative can be assessed.

Table 5: Sensitivity to the distillation weight.

λ	EER (%)	EER + TTA (%)
0.5	1.440 ± 0.192	1.354 ± 0.200
1.0 (adopted)	1.334 ± 0.097	1.242 ± 0.074
2.0	1.273 ± 0.167	1.185 ± 0.185

EER = equal error rate; TTA = test-time augmentation; λ = distillation weight.

Table 6 compares the ResNet-50 teacher with a ConvNeXt-Tiny teacher trained under the identical pipeline and distilled into the same student. ConvNeXt-Tiny is the weaker teacher on this database (1.370% versus 1.145%), and its distilled student is correspondingly weaker (1.371% versus 1.242%). The more recent architecture therefore does not provide better embeddings for this database. ConvNeXt was designed and evaluated on large image collections such as ImageNet, and a possible explanation is that 1,600 training identities per fold are too few for this architecture to reach its full accuracy; the present experiments do not test this explanation. The result supports the choice of ResNet-50 as the teacher. Table 6 also shows that the fraction of the gap closed is misleading on its own, because distillation from the ConvNeXt teacher closes 95.0% of a much narrower gap yet yields a worse student. The accuracy of the distilled student therefore depends more on the absolute error rate of the teacher than on the fraction of the student–teacher gap that distillation closes. The student is EfficientNet-B0 in all rows of table 6, and the values are five-fold means obtained with TTA.

Table 6: Comparison of the two teacher architectures.

Teacher	Params (M)	Teacher EER (%)	Distilled student EER (%)	Gap closed (%)
ResNet-50	25.61	1.145	1.242	60.4
ConvNeXt-Tiny	≈28.6	1.370	1.371	95.0
— (no distillation)	—	—	1.390	—

EER = equal error rate; M = million.

4.4. Computational Efficiency

Table 7 reports the four efficiency measures for the student and its teacher. The distilled EfficientNet-B0 requires 5.08M parameters, 1.08 GFLOPs and 20.3 MB, against 25.61M, 10.8 GFLOPs and 102.4 MB for ResNet-50, which corresponds to one fifth of the parameters and one tenth of the operations, and it runs 2.4 times faster on the CPU (42.35 versus 102.81 ms per image). On the GPU, however, ResNet-50 is faster (6.35 versus 8.27 ms), likely because its dense convolutions are highly optimized by the NVIDIA CUDA Deep Neural Network library (cuDNN), whereas the depthwise convolutions of EfficientNet are not. The efficiency advantage of the proposed model is therefore stated for the central processing unit of the evaluation host and not for graphics-processing-unit inference. No constrained or embedded device was used, and the quoted latencies cover the embedding network alone; the cost of segmentation and of the seven-view TTA is excluded. Distillation adds no inference cost, because the distilled and baseline students are architecturally identical. The measurements in table 7 use a single 256×256 input in FP32, and latency is the mean time per image at batch size one.

Table 7: Efficiency of the student and the teacher.

Model	Params (M)	GFLOPs	Size (MB)	GPU latency (ms)	CPU latency (ms)
EfficientNet-B0 (student, distilled)	5.08	1.08	20.3	8.27	42.35
ResNet-50 (teacher)	25.61	10.8	102.4	6.35	102.81

CPU = central processing unit; GFLOPs = giga floating-point operations; GPU = graphics processing unit; M = million; MB = megabytes.

4.5. Robustness to Mask Quality and Segmentation Failures

The occlusion-aware attention module takes the segmentation mask as an input, so the accuracy of the distilled model may depend on the quality of that mask. To measure this dependence, the mask was degraded at inference while the weights of the distilled model were kept fixed, and table 8 gives the results. Dilating or eroding the mask by up to five pixels, or corrupting up to 30% of its pixels,

changes the EER by at most 0.04 percentage points, and removing the mask entirely raises it only from 1.334% to 1.388%. The model is therefore robust to the quality of the mask. This result is consistent with the ablation results in table 4, because the segmentation and normalization stage already set most occluded pixels to zero before feature extraction, which leaves the attention module little residual occlusion to use. The degradations in table 8 are applied at inference only, and the values are five-fold means without TTA.

The 46 images that could not be segmented are treated as failure-to-acquire cases. Their worst-case effect on the error estimate is bounded by counting every failed acquisition as a recognition error, and under this conservative convention the EER of the distilled model rises from 1.242% to 1.469%. This is an upper bound, and because the same failure rate applies to every model, the ranking of the models is unaffected.

Table 8: Sensitivity of the distilled model to occlusion-mask quality.

Occlusion-mask condition	EER (%)
Original mask	1.334 ± 0.097
Dilated by 2 px	1.324 ± 0.099
Dilated by 5 px	1.320 ± 0.099
Eroded by 2 px	1.349 ± 0.103
Eroded by 5 px	1.360 ± 0.104
Random corruption, 5%	1.339 ± 0.096
Random corruption, 15%	1.348 ± 0.100
Random corruption, 30%	1.374 ± 0.108
Mask removed	1.388 ± 0.114

EER = equal error rate.

4.6. Validity Analysis of the Input Region

To test whether full-frame evaluation on CASIA-Iris-Thousand measures iris matching, the same EfficientNet-B0 was trained on four input regions under identical folds, seeds, pair sampling and schedule, so that the input region is the only variable (table 9, figure 8). All four configurations were generated with the final pipeline; no figure in this analysis derives from a preliminary configuration.

The full eye frame yields an EER of 0.457% ± 0.087%, better than any iris-only configuration reported in this paper. Erasing the iris and the pupil entirely, however, still leaves 0.571% ± 0.154%, so the periocular region alone recovers almost all of that performance, and restoring the whole iris improves the error by only 0.114 percentage points. The coarse iris crop, whose square bounding box retains surrounding tissue in its corners, is intermediate at 1.186% ± 0.190%, whereas the segmented and normalized iris, which is the only configuration from which non-iris content has been removed, yields 1.475% ± 0.076%. The error therefore rises monotonically as non-iris information is progressively withdrawn, which is the signature of a confound rather than of a progressively harder task.

These results indicate that full-frame evaluation is not a suitable basis for comparing iris recognition systems on this database, because an error rate below 1% obtained from full-frame input mainly measures periocular appearance rather than iris texture. Among the configurations examined, the segmented and normalized iris is the only one in which the measured quantity is the iris itself. For the same reason, the 1.242% attained by the distilled student is obtained under a more restrictive input condition than the 0.457% of the full frame, and the two are not comparable. In addition, three of the five full-frame folds and one periocular fold triggered the leakage guard (an EER below 0.5%) although the partitioning is subject-disjoint. In these cases, the guard responds to periocular confounding rather than to identity leakage, which is the behavior it was designed to detect.

Table 9: Equal error rate obtained from four input regions.

Input region	EER (%)	Interpretation
Full eye frame	0.457 ± 0.087	Non-iris content removed, Iris-only measurement, Recovers almost all of the full-frame performance, without the iris.
Periocular only (iris and pupil erased)	0.571 ± 0.154	
Coarse iris crop (unnormalized)	1.186 ± 0.190	Square crop still retains surrounding tissue
Segmented and normalized iris	**1.475 ± 0.076**	

EER = equal error rate.

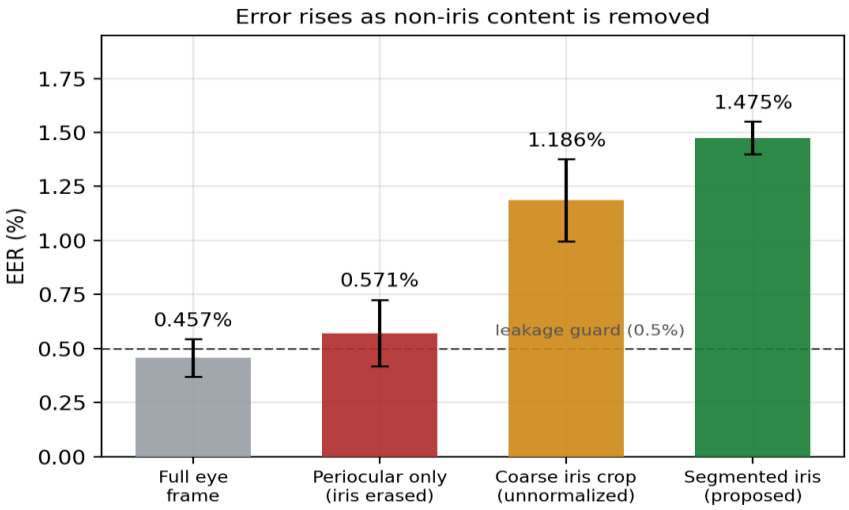


Figure 8: Equal error rate obtained from four input regions.

4.7. Cross-Fold Stability

Figure 9 reports per-fold error rates, and table 10 lists them for each model. Performance varies across folds, and for DenseNet-201 a single split could have given any EER between 1.100% and 1.515%, which shows why results over several folds are needed. This variation reflects differences in the difficulty of the subject-disjoint partitions rather than instability of the recognizer, because the distilled model's cross-fold standard deviation is 0.074 percentage points (a coefficient of variation of about 6%), lower than that of the teacher (0.088) and of DenseNet-201 (0.135), and TTA reduces it below 0.097. The per-fold values in table 10 are obtained with TTA.

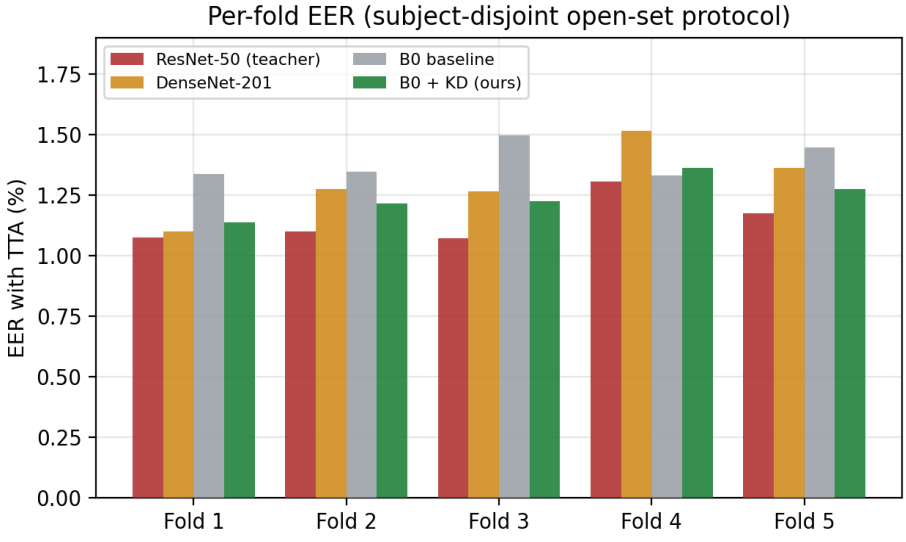


Figure 9: Equal error rate of each fold for the main models.

Fold 4 is the hardest fold for ResNet-50, DenseNet-201 and the distilled model, but not for the EfficientNet-B0 baseline, for which it is the easiest; the difficulty is therefore partly model-dependent. Occlusion does not explain this difficulty, because the mean occlusion fraction ranges only from 16.1% to 16.7% across folds. No single acquisition factor was found to account for the per-fold differences, and they are reported as a property of the subject partitions. Table 10 does show one consistent pattern: Fold 4 is where the teacher is weakest, which explains why distillation brings no gain there.

Table 10: Equal error rate of each fold and the change produced by distillation.

Fold	ResNet-50 (teacher)	DenseNet-201	B0 baseline	B0 + KD	Change from KD (pp)
1	1.075	1.100	1.335	1.135	−0.200
2	1.100	1.275	1.345	1.215	−0.130
3	1.070	1.265	1.495	1.225	−0.270
4	1.305	1.515	1.330	1.360	+0.030
5	1.175	1.360	1.445	1.275	−0.170

KD = knowledge distillation; pp = percentage points.

5. Discussion

Within a segmentation-first pipeline free of identity leakage and periocular confounding, knowledge distillation closes 60.4% of the error gap between a compact model and a large teacher, achieving a 1.242% EER with 5.08M parameters. The results do not support two expectations that motivated the design, namely that a lightweight architecture would outperform the larger networks and that the attention modules would be a principal source of accuracy.

Absolute error rates reported by other work on this database are not directly comparable with those reported here, because they differ in input region (full-eye or periocular versus iris-only), in segmentation, in image selection, and in the partitioning of subjects versus images. The validity analysis reported here quantifies how large such differences can be, since on this database the choice of input region moves the EER by more than a factor of three, which exceeds every architectural effect measured in this paper. The only controlled comparison in this study is therefore the comparison of the five backbones in table 3, which were all trained and evaluated under the same protocol, and comparisons with error rates reported in other studies should be treated as indicative only.

Table 11 compares the results of this study with those of the studies reviewed in the related work section. Because these works differ in task, database, input region and protocol, the comparison shows their methodological differences rather than a ranking of absolute error rates, and the controlled comparison remains the one in table 3. In table 11, a dagger marks studies evaluated on different databases, which are listed for methodological context rather than for numerical comparison, and NR denotes a source that does not report an EER in a comparable form. Studies without a recognition EER, namely the segmentation study of Li *et al.* [15] and the presentation-attack study of Gautam *et al.* [19], are not included in table 11 because neither reports a recognition result; both are summarized in table 1 and discussed in the related works section.

Table 11: Comparison with related works.

Ref.	Database	Input	Method	KD	Protocol	EER (%)
[12]	CASIA v4 Dist.; MMU V2	Full-eye	Log-Gabor (non-DL); Hamming	No	Image-level split	NR (accuracy 87.68 / 94.47)
[18] †	CASIA v4 Dist.; UBIPr; AR	Iris + periocular	CNN; softmax+center; attention	No	Not subject-disjoint	NR on this benchmark
[26]	CASIA-Iris-Thousand	Iris only	MobileNetV3-L; triplet; distilled from ResNet-50	Yes	Single subject-level 70:15:15 split	1.409
This work	CASIA-Iris-Thousand	Iris only	EffNet-B0 5.08M; sub-center ArcFace; distilled from ResNet-50	Yes	Open-set, subject-disjoint, 5-fold	1.242

AR = Aleix and Robert face database; CASIA = Institute of Automation of the Chinese Academy of Sciences Institute of Automation; CNN = convolutional neural network; Dist. = Distance; DL = deep learning; EER = equal error rate; EffNet = EfficientNet; KD = knowledge distillation; MMU = Multimedia University; NR = not reported; UBIPr = University of Beira Interior Periocular database; † = evaluated on a different database.

All results are obtained within CASIA-Iris-Thousand. Cross-sensor and cross-database generalization, which is the main question for forensic deployment across agencies and acquisition devices, is therefore not established. The principal technical obstacle is that the learned segmentation front-end is calibrated to the optics and resolution of this sensor, so transfer to databases such as the IIT Delhi (IITD) iris database [50], ND-IRIS-0405 [51] or mobile near-infrared corpora requires either a sensor-agnostic segmenter or re-calibration, and a failure of that stage would be indistinguishable from a failure of the recognizer. Establishing cross-sensor performance under the same protocol, free of identity leakage and periocular confounding, with segmentation quality reported separately from matching accuracy, is the most important direction for future work.

The periocular experiment in the validity analysis uses coarse region masking rather than an exact anatomical decomposition. It suffices to demonstrate that substantial non-iris signal is present and exploited, but the figures should be read as evidence of contamination rather than an exact apportionment of information between regions.

About 0.23% of images (46 of 20,000) fail segmentation, predominantly acquisitions with severe eyeglass reflections. They are treated as failure-to-acquire cases and excluded from matching. Because the failures are non-random, their exclusion could lower the reported error. Counting every failed acquisition as an error raises the EER of the distilled model by at most 0.23 percentage points, from 1.242% to 1.469%.

The distilled student's accuracy is bounded by its teacher. Only two convolutional teacher architectures were compared. Transformer-based teachers, such as the Vision Transformer (ViT) [52] and the Swin Transformer [53], were not trained, because transformers generally need much larger training sets than the 1,600 identities available in each fold. Ensemble teachers, intermediate-layer distillation and relational distillation objectives were also not examined. In addition, all backbones share one set of hyperparameters; architecture-specific tuning, which could change the relative ranking of EfficientNet-B0 and B1, was not performed.

The attention modules that account for frequency and occlusion did not achieve the improvements they were intended to provide (Table 4). They are retained in the architecture for consistency with the reported experiments, but their contribution is small and arises mainly from their joint presence rather than from either module alone.

In summary, this study has four principal strengths. The evaluation protocol is open-set, subject-disjoint, and five-fold, and both eyes of each subject are assigned to the same partition to prevent identity leakage. The iris-only input eliminates periocular confounding, and the validity analysis quantifies the magnitude of this confound directly. Five backbone architectures are compared under an identical training and evaluation protocol. Efficiency is assessed with four complementary measures obtained under a common protocol rather than with parameter count alone. The study also has several limitations. All results are confined to a single database and sensor, which limits claims of generalization. Efficiency was measured on the CPU of the evaluation host rather than on a resource-constrained or embedded device. The 0.23% of images that failed segmentation were excluded from matching. The periocular analysis relied on coarse region masking and therefore provides only an approximate estimate of periocular influence. Only two convolutional teachers and a single shared set of hyperparameters were examined. Finally, the attention modules yielded no clear improvements in accuracy.

6. Conclusions

This study has presented a segmentation-first pipeline for iris recognition on CASIA-Iris-Thousand in which identity leakage and periocular confounding are controlled separately, and has used it to establish three findings. First, full-frame evaluation on this database measures mainly non-iris appearance, because erasing the iris and the pupil leaves the error almost unchanged, whereas the error rises monotonically as non-iris content is withdrawn. Full-frame protocols are therefore not a sound basis for comparing iris matchers. Second, under a subject-disjoint, open-set five-fold protocol the two largest networks achieved a lower EER than the compact networks, although error did not fall monotonically with model size, since the larger of the two EfficientNet students performed worse than

the smaller one; network family and capacity are therefore confounded in this comparison. Neither added student capacity, nor a more recent teacher architecture, nor the attention modules improved a compact matcher materially. Third, knowledge distillation reduced the error of the compact student at an unchanged parameter count and was the only intervention examined that did so. The evaluation is confined to one database and one sensor, and cross-sensor and cross-database generalization remain untested. Future work will extend the protocol to cross-sensor and cross-database settings and examine richer distillation objectives.

Author Contributions: Khalid Ibrahim Khalaf Jajan: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Wafaa Mustafa Abduallah:** Conceptualization, Supervision, Project administration, Writing – review & editing.

Data availability: The data generated during and/or analyzed during the current study are available in the Kaggle repository: https://www.kaggle.com/datasets/sondosaabed/casia-iris-thousand.

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

Funding: The authors did not receive support from any organization for the conducting of the study.

References

[1] K. Nguyen, H. Proença, and F. Alonso-Fernandez, "Deep learning for iris recognition: a survey," *ACM Computing Surveys*, vol. 56, no. 9, art. 223, pp. 1–35, 2024, doi: 10.1145/3651306.

[2] Y. Yan, Q. Wang, H. Zhu, and W. Jiang, "Iris-LAHNet: a lightweight attention-guided high-resolution network for iris segmentation and localization," *Multimedia Systems*, vol. 30, no. 2, art. 85, 2024, doi: 10.1007/s00530-024-01280-5.

[3] S. Bhatt, J. S. Sehrawat, and V. Gupta, "A systematic review of iris biometrics in forensic science: applications and challenges," *Egyptian Journal of Forensic Sciences*, vol. 15, art. 12, 2025, doi: 10.1186/s41935-025-00431-7.

[4] Y. Yin, S. He, R. Zhang, H. Chang, and J. Zhang, "Deep learning for iris recognition: a review," *Neural Computing and Applications*, vol. 37, no. 17, pp. 11125–11173, 2025, doi: 10.1007/s00521-025-11109-5.

[5] M. R. Sumi, P. Das, A. Hossain, S. Dey, and S. Schuckers, "A comprehensive evaluation of iris segmentation on benchmarking datasets," *Sensors*, vol. 24, no. 21, art. 7079, 2024, doi: 10.3390/s24217079.

[6] H. Proença and J. C. Neves, "Deep-PRWIS: periocular recognition without the iris and sclera using deep learning frameworks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 888–896, 2018, doi: 10.1109/TIFS.2017.2771230.

[7] D. P. Benalcazar, J. E. Tapia, M. Vasquez, L. Causa, E. L. Droguett, and C. Busch, "Toward an efficient iris recognition system on embedded devices," *IEEE Access*, vol. 11, pp. 133577–133590, 2023, doi: 10.1109/ACCESS.2023.3337033.

[8] E. Caldeira, P. C. Neto, M. Huber, N. Damer, and A. F. Sequeira, "Model compression techniques in biometrics applications: a survey," *Information Fusion*, vol. 114, art. 102657, 2025, doi: 10.1016/j.inffus.2024.102657.

[9] J. J. Winston, D. J. Hemanth, A. Angelopoulou, and E. Kapetanios, "Hybrid deep convolutional neural models for iris image recognition," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 9481–9503, 2022, doi: 10.1007/s11042-021-11482-y.

[10] J. E. Zambrano, D. P. Benalcazar, C. A. Perez, and K. W. Bowyer, "Iris recognition using low-level CNN layers without training and single matching," *IEEE Access*, vol. 10, pp. 41276–41286, 2022, doi: 10.1109/ACCESS.2022.3166910.

[11] N. Ahmad, K. S. Yadav, A. M. Kirupakaran, S. A. Barlaskar, and R. H. Laskar, "Design and development of an integrated approach towards detection and tracking of iris using deep learning," *Multimedia Tools and Applications*, vol. 83, pp. 44661–44683, 2024, doi: 10.1007/s11042-023-17433-z.

[12] A. K. Nsaif, S. H. M. Ali, A. K. Nseaf, K. N. Jassim, A. Al-Qaraghuli, and R. Sulaiman, "Robust and swift iris recognition at distance based on novel pupil segmentation," *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 10, pp. 9184–9206, 2022, doi: 10.1016/j.jksuci.2022.09.002.

[13] Y. Sun, Y. Lu, Y. Liu, S. Liu, and Z. Liu, "LiSeNet: multitask lightweight segmentation network for accurate and complete iris segmentation," *Journal of Electronic Imaging*, vol. 31, no. 5, art. 053035, 2022, doi: 10.1117/1.JEI.31.5.053035.

[14] V. I. D'Alessandro, L. De Palma, F. Attivissimo, A. Di Nisio, and A. M. L. Lanzolla, "U-Net convolutional neural network for multisource heterogeneous iris segmentation," *IEEE International Symposium on Medical Measurements and Applications* (MeMeA), Jeju, Republic of Korea, 2023, pp. 1–5, doi: 10.1109/MeMeA57477.2023.10171896.

[15] Y. Li, N. Chen, G. Zhao, and Y. Shen, "KD-Eye: lightweight pupil segmentation for eye tracking on VR headsets via knowledge distillation," *International Conference on Wireless Artificial Intelligent Computing Systems and Applications* (WASA), Qingdao, China, 2024, Lecture Notes in Computer Science, vol. 14997, pp. 209–220, doi: 10.1007/978-3-031-71464-1_18.

[16] S. A. Latif, K. A. Sidek, E. A. Bakar, and A. H. A. Hashim, "Online multimodal compression using pruning and knowledge distillation for iris recognition," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 37, no. 2, pp. 68–81, 2024, doi: 10.37934/araset.37.2.6881.

[17] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: additive angular margin loss for deep face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 5962–5979, 2022, doi: 10.1109/TPAMI.2021.3087709.

[18] Z. Luo, J. Li, and Y. Zhu, "A deep feature fusion network based on multiple attention mechanisms for joint iris–periocular biometric recognition," *IEEE Signal Processing Letters*, vol. 28, pp. 1060–1064, 2021, doi: 10.1109/LSP.2021.3079850.

[19] G. Gautam, A. Raj, and S. Mukhopadhyay, "Deep supervised class encoding for iris presentation attack detection," *Digital Signal Processing*, vol. 121, art. 103329, 2022, doi: 10.1016/j.dsp.2021.103329.

[20] J. Wei, H. Huang, Y. Wang, R. He, and Z. Sun, "Towards more discriminative and robust iris recognition by learning uncertain factors," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 865–879, 2022, doi: 10.1109/TIFS.2022.3154240.

[21] J. Wei, Y. Wang, H. Huang, R. He, Z. Sun, and X. Gao, "Contextual measures for iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 57–70, 2023, doi: 10.1109/TIFS.2022.3221897.

[22] J. Wei, Y. Wang, Y. Li, R. He, and Z. Sun, "Cross-spectral iris recognition by learning device-specific band," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 3810–3824, 2022, doi: 10.1109/TCSVT.2021.3117291.

[23] L. Jia, X. Shi, Q. Sun, X. Tang, and P. Li, "Second-order convolutional networks for iris recognition," *Applied Intelligence*, vol. 52, no. 10, pp. 11273–11287, 2022, doi: 10.1007/s10489-021-02925-y.

[24] X. Sun, C. Wang, Y. Wang, J. Wei, and Z. Sun, "IrisFormer: a dedicated transformer framework for iris recognition," *IEEE Signal Processing Letters*, vol. 32, pp. 431–435, 2025, doi: 10.1109/LSP.2024.3522856.

[25] Y. Luo, X. Zhao, H. Shen, C. Fu, and H. Jiang, "A novel deep learning network for iris recognition using non-global interactive channel attention and lightweight spatial attention," *Signal, Image and Video Processing*, vol. 19, art. 950, 2025, doi: 10.1007/s11760-025-04561-2.

[26] T.-T. Huynh, D.-T. Huynh, D.-T. Phu, H.-S. Nguyen, and Q. H. Nguyen, "MobiIris: attention-enhanced lightweight iris recognition with knowledge distillation and quantization," *Computers, Materials & Continua, vol. 87*, no. 3, art. 19, 2026, doi: 10.32604/cmc.2026.076623.

[27] S. Ramesh and V. Krishnaveni, "Transformer-based iris verification with attention-guided segmentation and Siamese learning," *Scientific Reports*, vol. 16, art. 22204, 2026, doi: 10.1038/s41598-026-53259-z.

[28] K. Lu, R. Wu, and E. Bao, "Multilevel parallel attention knowledge distillation for multimodal biometric recognition," *Engineering Applications of Artificial Intelligence*, vol. 154, art. 110865, 2025, doi: 10.1016/j.engappai.2025.110865.

[29] C. Liu, L. Yang, W. Zhou, Y. Li, and F. Hao, "Deep distillation hashing for palmprint and finger vein retrieval," *IET Biometrics*, vol. 2025, art. 9017371, 2025, doi: 10.1049/bme2/9017371.

[30] Y. Liu, W. Shen, D. Wu, and J. Shao, "IrisST-Net for iris segmentation and contour parameters extraction," *Applied Intelligence*, vol. 53, no. 9, pp. 11267–11281, 2023, doi: 10.1007/s10489-022-03973-8.

[31] J. Muhammad, C. Wang, Y. Wang, K. Zhang, and Z. Sun, "Iris-GuideNet: guided localisation and segmentation network for unconstrained iris biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2723–2736, 2023, doi: 10.1109/TIFS.2023.3268504.

[32] J. Li and X. Feng, "Double-center-based iris localization and segmentation in cooperative environment with visible illumination," Sensors, vol. 23, no. 4, art. 2238, 2023, doi: 10.3390/s23042238.

[33] G. Huo, D. Lin, and M. Yuan, "Iris segmentation method based on improved UNet++," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 41249–41269, 2022, doi: 10.1007/s11042-022-13198-z.

[34] J. Sun, S. Zhao, S. Miao, X. Wang, and Y. Yu, "Open-set iris recognition based on deep learning," *IET Image Processing*, vol. 16, no. 9, pp. 2361–2372, 2022, doi: 10.1049/ipr2.12493.

[35] J. Sun, S. Zhao, Y. Yu, X. Wang, and L. Zhou, "Iris recognition based on local circular Gabor filters and multi-scale convolution feature fusion network," *Multimedia Tools and Applications*, vol. 81, no. 23, pp. 33051–33065, 2022, doi: 10.1007/s11042-022-13098-2.

[36] A. Rubio and B. Magnier, "Preprocessing of iris images for BSIF-based biometric systems: binary detected edges and iris unwrapping," *Sensors*, vol. 24, no. 15, art. 4805, 2024, doi: 10.3390/s24154805.

[37] A. Sharma, D. Kumar, and G. Gupta, "Enhanced iris recognition with histogram cut selection and genetic algorithms for robust classification," *Procedia Computer Science*, vol. 258, pp. 2846–2859, 2025, doi: 10.1016/j.procs.2025.04.545.

[38] M. G. Aliyu, S. Jamel, and M. Danlami, "An enhanced iris recognition using deep learning and 2D Chebyshev wavelet filter," *Journal of Information and Communication Technology*, vol. 25, no. 3, pp. 98–123, 2026, doi: 10.32890/jict2026.25.3.5.

[39] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004, doi: 10.1109/TCSVT.2003.818350.

[40] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," *European Conference on Computer Vision* (ECCV), Munich, Germany, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[41] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1655–1668, 2019, doi: 10.1109/TPAMI.2018.2846566.

[42] J. Deng, J. Guo, T. Liu, M. Gong, and S. Zafeiriou, "Sub-center ArcFace: boosting face recognition by large-scale noisy web faces," in Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI, *Lecture Notes in Computer Science*, vol. 12356. Cham, Switzerland: Springer, 2020, pp. 741–757, doi: 10.1007/978-3-030-58621-8_43.

[43] Institute of Automation, Chinese Academy of Sciences, "CASIA Iris Image Database Version 4.0 (CASIA-IrisV4)," 2010. [Online]. Available: http://biometrics.idealtest.org/.

[44] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: a nested U-Net architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA/ML-CDS), held with MICCAI, Granada, Spain, 2018, *Lecture Notes in Computer Science*, vol. 11045, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.

[45] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), Salt Lake City, UT, USA, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[46] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," *International Conference on Machine Learning* (ICML), Long Beach, CA, USA, 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html.

[47] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), New Orleans, LA, USA, 2022, pp. 11976–11986, doi: 10.1109/CVPR52688.2022.01167.

[48] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *International Conference on Learning Representations* (ICLR), New Orleans, LA, USA, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7.

[49] I. Loshchilov and F. Hutter, "SGDR: stochastic gradient descent with warm restarts," *International Conference on Learning Representations* (ICLR), Toulon, France, 2017. [Online]. Available: https://openreview.net/forum?id=Skq89Scxx.

[50] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," Pattern Recognition, vol. 43, no. 3, pp. 1016–1026, 2010, doi: 10.1016/j.patcog.2009.08.016.

[51] P. J. Phillips *et al*., "FRVT 2006 and ICE 2006 large-scale experimental results," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pp. 831–846, 2010, doi: 10.1109/TPAMI.2009.59

[52] A. Dosovitskiy *et al*., "An image is worth 16x16 words: transformers for image recognition at scale," International Conference on Learning Representations (ICLR), Vienna, Austria (virtual), 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy.

[53] Z. Liu *et al*., "Swin Transformer: hierarchical vision transformer using shifted windows," *IEEE/CVF International Conference on Computer Vision* (ICCV), Montreal, QC, Canada, 2021, pp. 10012–10022, doi: 10.1109/ICCV48922.2021.00986.