



ARTICLE

Efficient Iris Recognition via Polar Representation and Radial Stripe Attention

Trong-Thua Huynh^{1,*}, De-Thu Huynh², Cong-Sang Duong¹, Hong-Son Nguyen¹, Quoc H. Nguyen³
and Lam-Thanh Tu⁴

¹Faculty of Information Technology II, Posts and Telecommunications Institute of Technology, 11 Nguyen Dinh Chieu Street, Sai Gon Ward, Ho Chi Minh City, Viet Nam

²School of Computer Science & Engineering, The Saigon International University, 16 Tong Huu Dinh Street, An Khanh Ward, Ho Chi Minh City, Viet Nam

³Institute of Digital Technology, Thu Dau Mot University, 06 Tran Van On Street, Phu Loi Ward, Ho Chi Minh City, Viet Nam

⁴Advanced Intelligent Technology Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, 19 Nguyen Huu Tho Street, Tan Hung Ward, Ho Chi Minh City, Viet Nam

*Corresponding Author: Trong-Thua Huynh. Email: thuaht@ptit.edu.vn

Received: 13 February 2026; Accepted: 27 April 2026; Published: 27 May 2026

ABSTRACT: Deep iris recognition models are often trained on Cartesian grids, whereas iris texture follows a concentric structure with angular periodicity. This representational mismatch can weaken rotation robustness and limit pupil-to-limbus context modeling, while many pipelines still rely on accurate segmentation masks. We propose *RadialFormer*, an efficient mask-free iris recognition framework that performs representation learning directly in the polar domain. The pipeline first estimates pupil/iris parameters $(c_x, c_y, r_{in}, r_{out})$ using a percentile radial-gradient operator with anatomical constraints, and then applies a crop-based polar transform to obtain a compact 64×512 unwrapped iris map. To better match polar geometry, we introduce *Learnable Polar Position Encoding* (LPPE) with separable radial-angular embeddings, where Fourier terms in the angular branch enforce continuity at $\theta = 0/2\pi$. We further propose *Radial Stripe Window Attention* (RSWA), which computes self-attention within full-height radial stripes and uses modular angular shifting to preserve circular consistency. Trained end-to-end with batch-hard triplet loss under $P \times K$ sampling, RadialFormer achieves 99.04% TPR@1%FPR with 0.48% EER on CASIA-V4-Lamp, and 93.63% TPR@1%FPR with 2.92% EER on CASIA-V4-Interval. Ablation and cross-dataset evaluations further validate the contributions of polar processing, LPPE, and RSWA and demonstrate robust generalization across acquisition conditions. Under the same input resolution, RadialFormer reduces computation by about 3.5× compared with a standard transformer baseline while maintaining competitive recognition accuracy.

KEYWORDS: Iris recognition; polar unwrapping; vision transformer; positional encoding; window attention; metric learning

1 Introduction

Iris recognition is widely regarded as a highly reliable biometric modality because iris texture is highly distinctive and largely stable during adulthood, while acquisition is non-invasive [1]. It has therefore been deployed in large-scale scenarios such as border control, national identity systems, and mobile authentication [2]. Although classical iris-recognition pipelines have long incorporated rubber-sheet normalization, many recent learning-based systems still employ encoder designs inherited from generic Cartesian image modeling, which do not explicitly account for the iris' concentric anatomy and angular periodicity.

1.1 Limitations of Conventional Deep Iris Pipelines

Classical iris recognition follows a multi-stage pipeline popularized by Daugman [3], including iris delineation, rubber-sheet normalization, feature encoding, and matching. While effective under controlled acquisition, this decomposition becomes fragile in non-ideal imagery due to several factors:

- **Error propagation:** Localization errors propagate to normalization and feature extraction, often causing substantial degradation under unconstrained conditions [4].
- **Occlusions and reflections:** Eyelids, eyelashes, and specular highlights distort the appearance near boundaries and corrupt iris texture, making both boundary estimation and downstream feature learning less reliable [5].
- **Rotation handling overhead:** In-plane rotation typically requires explicit compensation in Cartesian space, introducing additional computation and potential alignment error.
- **Limited geometric adaptation:** Rectangular receptive fields and non-cyclic spatial modeling do not explicitly reflect the anisotropy between radial (pupil-to-limbus) and angular structures, which may underutilize iris topology.

Deep learning has improved individual stages of the pipeline. CNN-based segmenters enhance robustness to noise and occlusion [6,7], and learned feature extractors can outperform hand-crafted codes in many settings [8,9]. However, many approaches still rely on segmentation-dependent multi-stage processing and encoder designs inherited from Cartesian image modeling, which increases system complexity and makes recognition performance more sensitive to localization and segmentation errors.

1.2 Why Polar Geometry Matters for Iris Representation Learning

The iris has an inherently polar organization: discriminative texture is organized around the pupil center, and the angular coordinate is periodic. This makes polar-domain representation learning attractive, where in-plane rotation is naturally converted into an (approximate) circular shift along the angular axis, and radial context can be modeled explicitly from pupil to limbus. A key challenge is to realize these benefits without introducing heavy preprocessing or segmentation dependency.

Vision transformers capture long-range dependencies via self-attention [10,11], and shifted-window designs improve computational efficiency [11]. Nevertheless, standard transformer formulations are typically inherited from generic Cartesian image modeling and do not explicitly reflect the radial–angular structure of polar iris maps: windowing is typically non-cyclic along the angular axis, positional encodings are designed for Cartesian coordinates, and many pipelines still assume segmented iris regions.

To address these issues, we propose *RadialFormer*, a polar-aware iris recognition framework that more explicitly aligns representation learning with iris geometry. Our approach consists of three main steps: (i) estimates pupil/iris parameters without pixel-wise segmentation masks, (ii) performs efficient crop-based polar unwrapping with angular wrap-around, and (iii) introduces geometry-aware transformer components—*Learnable Polar Position Encoding* (LPPE) and *Radial Stripe Window Attention* (RSWA)—to model angular periodicity and full radial context. Details are provided in [Section 3](#), with experimental results in [Section 4](#).

2 Related Work

2.1 Iris Recognition: Classical Pipelines and Deep Learning

Classical iris recognition is largely built on the pipeline popularized by Daugman [1,3], including iris delineation, rubber-sheet normalization, feature encoding (e.g., IrisCode), and matching, where rubber-sheet normalization established polar representation as a standard intermediate step for handling iris

annularity and rotation. Open-source systems such as OSIRIS have also provided reproducible implementations of classical iris recognition pipelines [12]. Subsequent studies improved delineation via edge/Hough search [5], geodesic active contours [13], level-set formulations [14], and refined integro-differential operators [15]. However, performance in non-ideal imagery remains strongly coupled with boundary quality: occlusions, specular highlights, blur, and illumination changes can distort pupil/limbus estimates and propagate errors into normalization and matching [4].

Deep learning has strengthened both segmentation and recognition. U-Net-like segmenters improve pixel-level masking robustness [6,7,16,17], and CNN-based recognition models learn discriminative iris embeddings from normalized strips or iris-centered crops [8,18,19], although the encoder backbone in such pipelines is often still adapted from generic image modeling and does not explicitly distinguish radial and angular positional structure. Yet, many pipelines remain multi-stage and segmentation-dependent; surveys note that in cross-sensor and unconstrained settings, segmentation/normalization errors often dominate failure modes [9], while more recent studies have also emphasized the importance of stronger representation learning, attention-based modeling, and loss design for improving robustness in practical iris recognition [20–22]. These observations motivate approaches that reduce reliance on pixel-wise masks while better aligning representation learning with iris geometry.

2.2 Transformers, Metric Learning, and Polar Geometry

Vision transformers model long-range dependencies via self-attention [10], while Swin Transformer improves scalability through local shifted-window attention [11]. Recent transformer-based iris studies have begun to explore this direction on normalized iris images or iris-centered inputs, but typical formulations still inherit assumptions from generic Cartesian image modeling that are not fully suited to polar iris data: rectangular, non-cyclic boundaries and generic positional encodings that do not explicitly account for radial–angular anisotropy or angular periodicity.

For open-set biometric verification, metric learning is widely adopted because it produces similarity-comparable embeddings without large classification heads. Triplet objectives (FaceNet [23]) and batch-hard mining [24] are commonly used, alongside related formulations such as lifted structured loss [25], N-pair loss [26], multi-similarity loss [27], and additive angular margin loss [28]. These objectives are attractive for iris verification because they directly optimize intra-class compactness and inter-class separation under cosine/Euclidean distance. In parallel, recent iris-recognition studies have explored attention-based formulations, uncertainty-aware representations, and stronger margin-based objectives to improve robustness under challenging acquisition conditions [20–22,29]. These developments reinforce the importance of representation design, but they do not explicitly address the radial–angular structure and angular periodicity of polar iris data. Polar parameterizations are natural for circular structures and have been explored in other domains with radial layouts [30,31]. In iris recognition, rubber-sheet normalization is standard [1] but is often treated as a fixed preprocessing step rather than a geometry-aware design principle inside the encoder. Our work integrates efficient crop-based polar unwrapping guided by mask-free parameter estimation, and designs polar-aware transformer components that explicitly encode angular periodicity and emphasize full pupil-to-limbus context modeling.

3 Method

3.1 Problem Formulation and Pipeline Overview

Let $I \in \mathbb{R}^{H \times W}$ be a grayscale eye image. Our goal is to learn a unit-normalized embedding function $f_{\Theta} : I \mapsto z \in \mathbb{R}^d$, $\|z\|_2 = 1$, so that samples from the same identity are close and different identities are well

separated. For verification we use cosine similarity $s(z_i, z_j) = z_i^\top z_j$, and for identification we perform nearest-neighbor search in \mathbb{R}^d .

Notation: We estimate iris parameters $(c_x, c_y, r_{in}, r_{out})$ (pupil center, inner/outer radii). Using these parameters we unwrap the iris annulus into a fixed-size polar map $I_{polar} \in \mathbb{R}^{1 \times H_{polar} \times W_{polar}}$, where H_{polar} and W_{polar} denote radial and angular resolutions. The angular coordinate θ is periodic on $[0, 2\pi)$.

Pipeline: RadialFormer is a segmentation-free iris recognition framework (Fig. 1) that performs representation learning directly in the polar domain: (i) mask-free localization to estimate $(c_x, c_y, r_{in}, r_{out})$; (ii) crop-based polar unwrapping to obtain I_{polar} ; (iii) a polar-aware encoder consisting of an asymmetric CNN stem, Learnable Polar Position Encoding (LPPE), and Radial Stripe Window Attention (RSWA); and (iv) metric learning using batch-hard triplet loss.

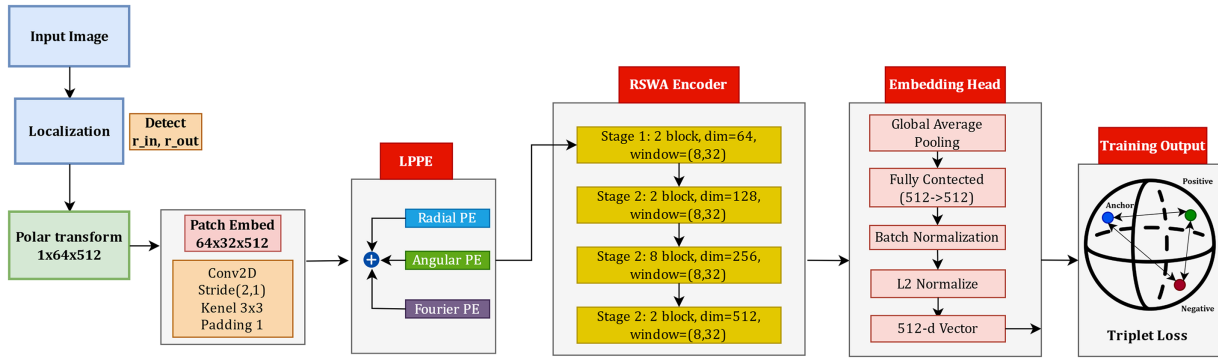


Figure 1: Overview of RadialFormer. PRG estimates $(c_x, c_y, r_{in}, r_{out})$, crop-based polar unwrapping produces I_{polar} with angular wrap-around, and a polar-aware encoder (CNN stem + LPPE + RSWA) outputs unit-norm embeddings optimized by metric learning.

Default setting: Unless stated otherwise, we use $(H_{polar}, W_{polar}) = (64, 512)$ and $d = 512$.

3.2 Mask-Free Localization via Percentile Radial-Gradient (PRG)

We localize the pupil and outer iris boundary without pixel-wise segmentation masks. The core idea is to score candidate circles using a percentile statistic of radial intensity change, which is robust to sparse angular outliers (specular highlights, eyelashes, partial eyelid occlusions). Algorithm 1 summarizes the complete coarse-to-fine PRG-based mask-free localization procedure used in this work.

Algorithm 1: Mask-free PRG localization (coarse-to-fine)

Input: image I ; thresholds τ_{hi}, σ ; ranges $[r_{in}^{\min}, r_{in}^{\max}]$, $[r_{out}^{\min}, r_{out}^{\max}]$; percentile τ ; center-prior weight β ; blend (λ, ρ) ; bounds $[\gamma_{min}, \gamma_{max}]$.

Output: $(\hat{c}_x, \hat{c}_y, \hat{r}_{in}, \hat{r}_{out})$.

- 1 $M_{spec} \leftarrow \mathbf{1}\{I \geq \tau_{hi}\}$; $\tilde{I} \leftarrow \text{Inpaint}(I, M_{spec})$; $\tilde{I}_\sigma \leftarrow \tilde{I} * \mathcal{N}(0, \sigma^2)$;
 - 2 Coarse search over (x_0, y_0) grid and $r \in [r_{in}^{\min}, r_{in}^{\max}]$ using $\tilde{\mathcal{R}}_\tau(r; x_0, y_0)$;
 - 3 Select $(\hat{c}_x, \hat{c}_y, \hat{r}_{in})$ by maximizing $\tilde{\mathcal{R}}_\tau(r; x_0, y_0) + \beta \phi_{center}(x_0, y_0)$; refine locally;
 - 4 $\tilde{r}_{out} \leftarrow \arg \max_{r \in [r_{out}^{\min}, r_{out}^{\max}]} \tilde{\mathcal{R}}_\tau(r; \hat{c}_x, \hat{c}_y)$;
 - 5 $\hat{r}_{out} \leftarrow (1 - \lambda)\tilde{r}_{out} + \lambda\rho\hat{r}_{in}$; clamp ratio to $[\gamma_{min}, \gamma_{max}]$;
 - 6 **return** $(\hat{c}_x, \hat{c}_y, \hat{r}_{in}, \hat{r}_{out})$;
-

Reflection-aware preprocessing: We suppress strong specular reflections by detecting saturated pixels and inpainting:

$$M_{\text{spec}}(x, y) = \mathbf{1}\{I(x, y) \geq \tau_{\text{hi}}\}, \quad \tilde{I} = \text{Inpaint}(I, M_{\text{spec}}), \quad (1)$$

followed by light Gaussian smoothing $\tilde{I}_\sigma = \tilde{I} * \mathcal{N}(0, \sigma^2)$, which stabilizes circle scoring in the presence of sensor noise and small specular artifacts.

Percentile radial-gradient score: For a candidate center (x_0, y_0) and radius r , define the radial gradient magnitude sampled along the circle:

$$\Gamma(r, \theta; x_0, y_0) = \left| \nabla \tilde{I}_\sigma(x_0 + r \cos \theta, y_0 + r \sin \theta) \cdot (\cos \theta, \sin \theta) \right|. \quad (2)$$

Instead of averaging over θ , we aggregate using a percentile:

$$\mathcal{R}_\tau(r; x_0, y_0) = \text{Percentile}_{\theta \in [0, 2\pi]}^\tau \Gamma(r, \theta; x_0, y_0), \quad \tau \in [50, 80]. \quad (3)$$

The percentile range $\tau \in [50, 80]$ is used to balance boundary sensitivity against robustness to sparse angular outliers. Lower percentiles are less selective for strong boundary evidence, whereas very high percentiles become increasingly sensitive to local artifacts such as specular highlights, eyelashes, and partial eyelid occlusion.

In practice we implement \mathcal{R}_τ with a finite-difference surrogate based on circular samples:

$$\widehat{\mathcal{R}}_\tau(r; x_0, y_0) = \text{Percentile}_\theta^\tau \left| \bar{I}(\mathcal{C}(r; x_0, y_0)) - \bar{I}(\mathcal{C}(r-1; x_0, y_0)) \right|, \quad (4)$$

where $\bar{I}(\mathcal{C}(r; \cdot))$ denotes bilinearly sampled intensities along a circle of radius r .

Coarse-to-fine pupil search: We search (x_0, y_0) on a coarse grid and then refine locally, using a coarse center step Δ_c (default $\Delta_c = 4$ pixels) followed by a small local refinement window around the best candidate. The pupil parameters are obtained by:

$$(\hat{c}_x, \hat{c}_y, \hat{r}_{\text{in}}) = \arg \max_{(x_0, y_0), r \in [r_{\text{in}}^{\text{min}}, r_{\text{in}}^{\text{max}}]} \widehat{\mathcal{R}}_\tau(r; x_0, y_0) + \beta \phi_{\text{center}}(x_0, y_0), \quad (5)$$

where ϕ_{center} is a weak center prior (e.g., distance to image center) and β controls its strength.

Outer boundary estimation with ratio regularization: With (\hat{c}_x, \hat{c}_y) fixed, we first estimate the raw limbus radius:

$$\tilde{r}_{\text{out}} = \arg \max_{r \in [r_{\text{out}}^{\text{min}}, r_{\text{out}}^{\text{max}}]} \widehat{\mathcal{R}}_\tau(r; \hat{c}_x, \hat{c}_y). \quad (6)$$

We then regularize using an anatomical ratio prior and clamp to plausible bounds, which helps suppress unstable outer-boundary estimates under weak limbus contrast, reflection, or partial occlusion:

$$\hat{r}_{\text{out}} = (1 - \lambda) \tilde{r}_{\text{out}} + \lambda \rho \hat{r}_{\text{in}}, \quad \frac{\hat{r}_{\text{out}}}{\hat{r}_{\text{in}}} \in [\gamma_{\text{min}}, \gamma_{\text{max}}]. \quad (7)$$

In practice, this ratio regularization restricts the outer-boundary search to anatomically plausible pupil-to-limbus proportions and reduces implausible solutions during inference. In our implementation, the outer-radius search interval is further constrained relative to the detected pupil radius, which improves stability across varying illumination conditions.

Implementation details for reproducibility: We sample each circle with M_θ uniform angles (default $M_\theta = W_{\text{polar}}$). Coarse grid step is Δ_c pixels (default $\Delta_c = 4$), and refinement searches a small neighborhood (e.g., ± 4 pixels, ± 3 radii). Ranges $[r_{\text{in}}^{\text{min}}, r_{\text{in}}^{\text{max}}]$ and $[r_{\text{out}}^{\text{min}}, r_{\text{out}}^{\text{max}}]$ are set per dataset using the training split statistics, so that the search space reflects the expected pupil and limbus scales of the corresponding acquisition setting rather than relying on a single universal range.

For clarity and reproducibility, Fig. 2 summarizes the complete preprocessing pipeline used before polar-domain representation learning, including reflection handling, PRG-based mask-free localization, iris-centered cropping, and crop-based polar transformation.

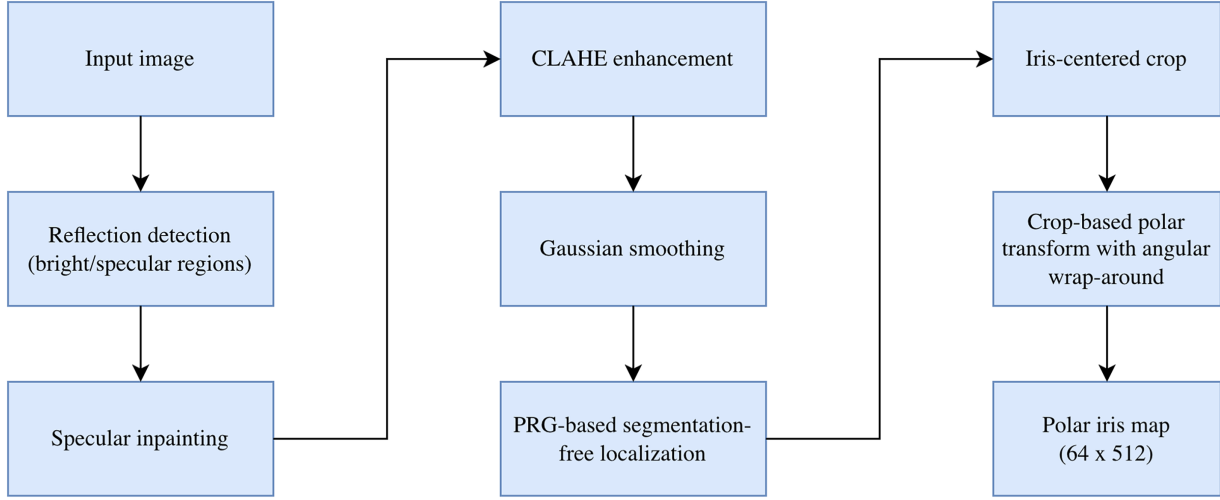


Figure 2: Preprocessing flow of RadialFormer. Starting from a grayscale eye image, the pipeline performs reflection detection and specular inpainting, contrast enhancement and smoothing, PRG-based mask-free localization, iris-centered cropping, and crop-based polar transformation with angular wrap-around, producing a fixed-size polar iris map for downstream representation learning.

In the implementation used for the present experiments, the preprocessing stage employs Telea inpainting for detected specular regions, CLAHE-based local contrast enhancement, light Gaussian smoothing, and a coarse-to-fine PRG search before crop-based polar unwrapping.

3.3 Crop-Based Polar Transformation

Given $(c_x, c_y, r_{\text{in}}, r_{\text{out}})$, we unwrap the iris annulus into a fixed-size polar map. Unlike a full-frame rubber-sheet transform, we compute the sampling grid *within an iris-centered crop*, which reduces background sampling, keeps the transform localized to the iris neighborhood, and improves computational efficiency (Fig. 3).

Iris-centered crop: We extract a square crop of side $2r_{\text{out}}$ centered at (c_x, c_y) :

$$I_{\text{crop}} = I[c_y - r_{\text{out}} : c_y + r_{\text{out}}, c_x - r_{\text{out}} : c_x + r_{\text{out}}], \quad (8)$$

where indices are clipped to the image bounds (out-of-range samples are handled by padding).

Sampling grid: We discretize radial and angular coordinates as:

$$r_i = r_{\text{in}} + \frac{i}{H_{\text{polar}} - 1} (r_{\text{out}} - r_{\text{in}}), \quad \theta_j = \frac{2\pi j}{W_{\text{polar}}}, \quad (9)$$

for $i \in \{0, \dots, H_{\text{polar}} - 1\}$ and $j \in \{0, \dots, W_{\text{polar}} - 1\}$.

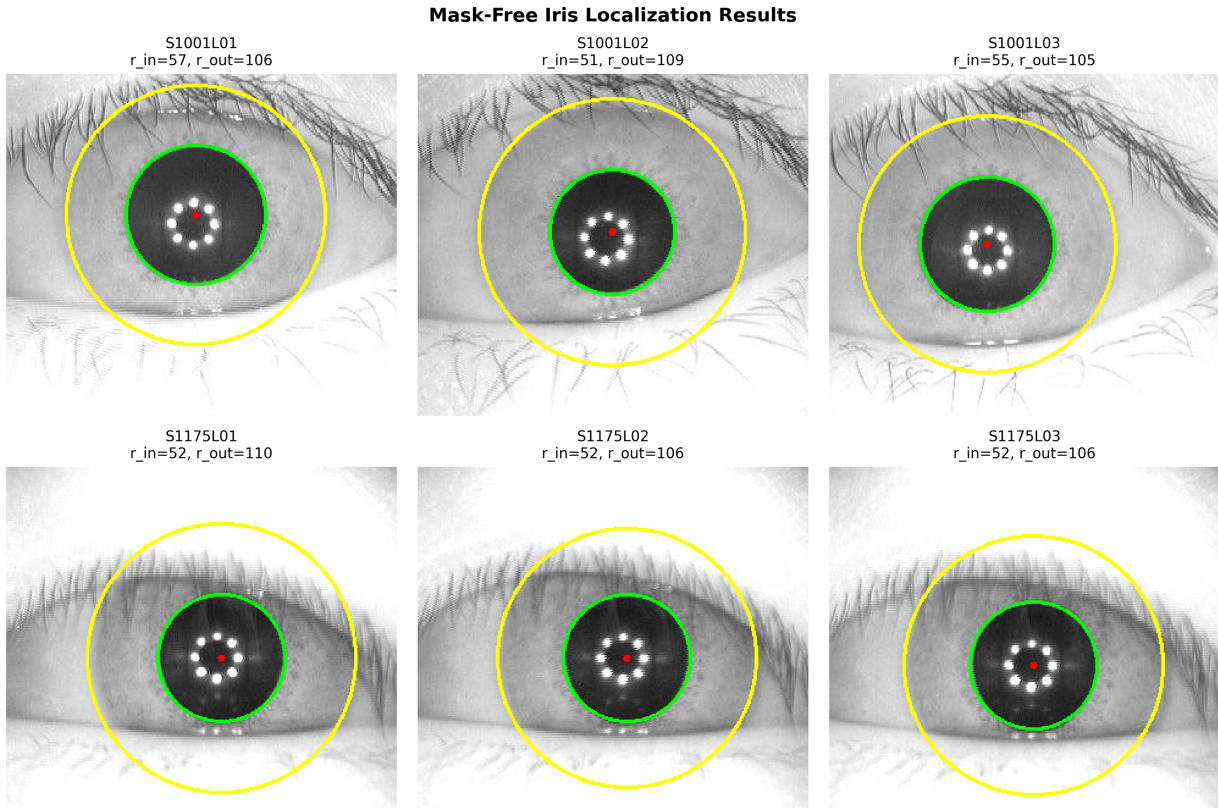


Figure 3: Qualitative PRG localization on CASIA-V4. Red: pupil center; green/blue: inner/outer boundaries. The method remains stable under specular highlights and partial occlusions without segmentation masks.

Radius-anchored mapping and modular wrap-around: The crop coordinate origin is $(r_{\text{out}}, r_{\text{out}})$, and under the shared-center approximation used in this work, the sampling locations are:

$$x_{i,j} = r_{\text{out}} + r_i \cos \theta_j, \quad y_{i,j} = r_{\text{out}} + r_i \sin \theta_j, \quad (10)$$

Shared-center approximation and practical motivation: Eq. (10) uses a shared-center approximation, i.e., the pupil and limbus are unwrapped with the same estimated center (c_x, c_y) . Classical rubber-sheet normalization may instead use a two-circle/non-concentric formulation to model pupil decentering more explicitly. In the present work, we adopt the shared-center approximation as a deliberate design choice because it yields a simpler and more efficient crop-based unwrapping procedure, reduces parameter sensitivity, and is well matched to the predominantly near-frontal NIR acquisition conditions of CASIA-IrisV4. In this setting, the approximation provides a favorable accuracy–efficiency trade-off while preserving stable recognition performance. A two-circle extension remains possible and may be beneficial for more strongly off-axis iris imagery, which we leave for future investigation.

The unwrapped map is obtained via bilinear interpolation:

$$I_{\text{polar}}[i, j] = I_{\text{crop}}(y_{i,j}, x_{i,j}). \quad (11)$$

To preserve angular periodicity, we implement *modular indexing* on the angular axis so that $\theta = 0$ and $\theta = 2\pi$ correspond to the same physical location.

Rotation-to-shift: A Cartesian in-plane rotation around (c_x, c_y) approximately becomes a circular shift along the angular axis of I_{polar} , up to interpolation error.

3.4 Polar-Aware Encoder: Asymmetric Stem, LPPE, and RSWA

The unwrapped map $I_{\text{polar}} \in \mathbb{R}^{1 \times H_r \times W_\theta}$ is anisotropic: fine-grained discriminative texture is mainly distributed along θ , while meaningful context spans pupil-to-limbus along r . Accordingly, we (i) preserve angular resolution, (ii) encode angular periodicity in positional representation, and (iii) align attention windows with the radial morphology.

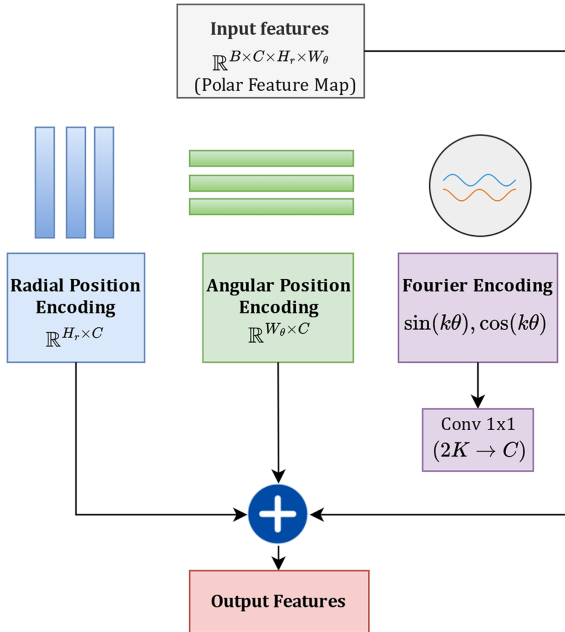
Asymmetric CNN stem (radial-only downsampling): We extract low-level features while downsampling only along r :

$$F_0 = \phi \left(\text{BN} \left(\text{Conv}_{3 \times 3}^{(s_r, 1)} (I_{\text{polar}}) \right) \right), \quad \phi = \text{GELU}, \quad (12)$$

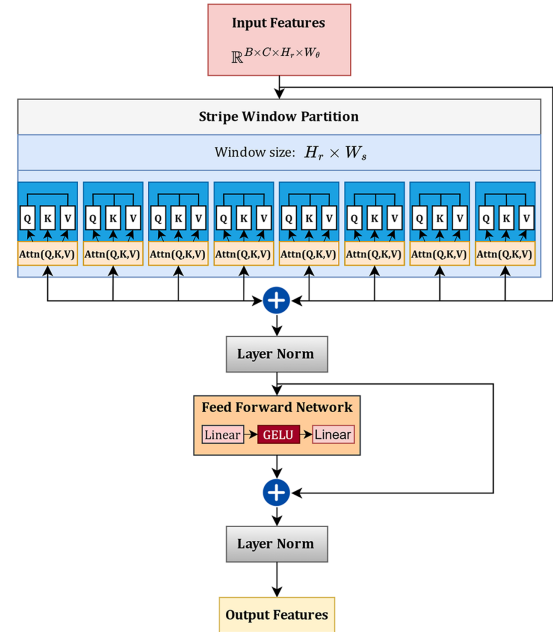
yielding $F_0 \in \mathbb{R}^{C \times H'_r \times W_\theta}$ with $H'_r = H_r/s_r$ and unchanged W_θ .

Learnable Polar Position Encoding (LPPE): Generic 2D positional encodings treat both axes as non-periodic (Fig. 4a). LPPE factorizes position into radial and angular embeddings and augments the angular branch with low-order Fourier features:

$$\text{PE}(i, j) = E_r(i) + E_\theta(j), \quad F_\theta(j) = \left[\sin(2\pi k j / W_\theta), \cos(2\pi k j / W_\theta) \right]_{k=1}^K. \quad (13)$$



a LPPE. Separable radial–angular embeddings with Fourier terms for θ periodicity.



b RSWA. Full-height radial stripes with modular angular shifting for circular consistency.

Figure 4: Geometry-aware encoder components of RadialFormer.

Fourier features are projected to C channels and injected additively:

$$\tilde{F}_0(:, i, j) = F_0(:, i, j) + \text{PE}(i, j) + \text{Proj}(F_\theta(j)). \quad (14)$$

Radial Stripe Window Attention (RSWA): To capture full radial dependencies without global attention, RSWA partitions \tilde{F}_0 into $S = \lfloor W_\theta/w_\theta \rfloor$ full-height stripes of size $(H'_r \times w_\theta)$ and applies multi-head self-attention within each stripe (Fig. 4b). We adopt a modular shifted strategy on θ to enable cross-stripe interaction while preserving circular consistency:

$$\tilde{F}(r, \theta) \mapsto \tilde{F}(r, (\theta + s) \bmod W_\theta), \quad s = \frac{w_\theta}{2}. \quad (15)$$

For each stripe, let $X \in \mathbb{R}^{(H'_r w_\theta) \times C}$ be the flattened tokens. We compute $Q = XW_Q$, $K = XW_K$, $V = XW_V$ and apply attention with polar-aware relative biases:

$$\text{Attn}(Q, K, V) = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} + B_r(\Delta r) + B_\theta(\Delta \theta) \right) V, \quad (16)$$

where Δr is the radial offset and $\Delta \theta$ is the angular offset inside a stripe (modulo w_θ).

Complexity: With stripe width w_θ , RSWA reduces global quadratic attention by restricting computation to stripes. The per-block complexity is $O(S (H'_r w_\theta)^2)$.

3.5 Embedding Head and Batch-Hard Triplet Learning

Embedding head: Given the final feature map $F_L \in \mathbb{R}^{C \times H \times W}$, we apply global average pooling and a linear projection:

$$h = \text{GAP}(F_L) \in \mathbb{R}^C, \quad \hat{z} = Wh + b \in \mathbb{R}^d, \quad z = \frac{\hat{z}}{\|\hat{z}\|_2}. \quad (17)$$

Batch-hard triplet loss: Each mini-batch samples P identities and K images per identity ($|\mathcal{B}| = PK$). For each anchor a , we select the hardest positive and hardest negative within the batch:

$$p^*(a) = \arg \max_{p: y_p = y_a, p \neq a} \|z_a - z_p\|_2, \quad n^*(a) = \arg \min_{n: y_n \neq y_a} \|z_a - z_n\|_2, \quad (18)$$

and optimize:

$$\mathcal{L}_{\text{tri}} = \frac{1}{|\mathcal{B}|} \sum_{a \in \mathcal{B}} \max \left(0, \|z_a - z_{p^*(a)}\|_2 - \|z_a - z_{n^*(a)}\|_2 + \alpha \right). \quad (19)$$

Inference: Given an input image, we compute $z = f_\Theta(I)$ and match using cosine similarity $z_i^\top z_j$.

4 Experiments

4.1 Datasets and Identity-Disjoint Splits

We evaluate RadialFormer on three near-infrared (NIR) benchmarks from CASIA-IrisV4: CASIA-V4-Interval (2639 images of 249 subjects, 320×280), collected under relatively controlled conditions; CASIA-V4-Lamp (16,212 images of 411 subjects), captured with lamp on/off illumination that induces strong specular highlights and contrast shifts; and CASIA-V4-Thousand (20,000+ images of 1000 subjects), which provides larger identity coverage and increased appearance diversity.

For clarity, Table 1 summarizes the main characteristics of the three CASIA-IrisV4 subsets used in this study.

Table 1: Comparison of the three CASIA-IrisV4 subsets used in this work.

Subset	Subjects	Images	Resolution	Acquisition Characteristics	Main Challenge
CASIA-V4-Interval	249	2639	320×280	Relatively controlled NIR capture	Moderate resolution and limited but non-negligible intra-class variation
CASIA-V4-Lamp	411	16,212	640×480	Lamp on/off illumination protocol	Strong specular highlights and contrast shifts
CASIA-V4-Thousand	1000	20,000+	640×480	Large-scale NIR capture with higher diversity	Larger identity coverage and stronger appearance variation

Identity-disjoint splits. For each dataset, we create subject-disjoint train/validation/test splits (70%/15%/15%), ensuring that no identity appears in more than one split. We evaluate mean \pm std over three random seeds, where the seed controls the subject partition and (for identification) the gallery selection.

Reproducibility note. All splits are generated at the identity level and kept fixed per seed across all methods to ensure fair comparisons.

4.2 Evaluation Protocols and Metrics

We evaluate verification and identification performance using standard biometric evaluation metrics. For verification, we evaluate Equal Error Rate (EER, \downarrow) and True Positive Rate (TPR, \uparrow) at fixed False Positive Rates (FPR) of 0.1% and 1%. For identification, we evaluate Rank-1 (R1) and Rank-5 (R5) accuracies.

Verification (open-set): For each test split, we compute unit-norm embeddings for all images and form (i) *genuine pairs* from all same-identity combinations and (ii) *impostor pairs* from different identities. Similarity is computed by cosine similarity $s_{ij} = z_i^T z_j$. We then compute the ROC curve and derive EER and TPR at the specified FPRs from the complete score sets. When the number of impostor pairs in a test split is insufficient to stably estimate the extreme operating point TPR@0.1% FPR, we omit it and report “-”.

Identification (1-to-N closed-set): For each test identity, we randomly select one image as the gallery and use the remaining images as probes. Each probe is matched against all gallery embeddings by cosine similarity, and R1/R5 are computed based on the top- k nearest neighbors. Gallery selection follows the random seed and is repeated across seeds.

4.3 Implementation Details

Preprocessing: All images are processed in grayscale at native resolution. We apply reflection-aware inpainting, perform mask-free localization, and unwrap the iris annulus into polar maps with $(H_{\text{polar}}, W_{\text{polar}}) = (64, 512)$ using modular wrap-around along θ .

Model configuration: The CNN stem downsamples only along the radial axis with stride (2, 1). Unless stated otherwise, RSWA uses stripe width $w_\theta = 16$, and the embedding dimension is $d = 512$.

Training protocol and fairness: We train with AdamW (weight decay 2×10^{-2}) and a cosine learning-rate schedule with a 5-epoch warmup to 3×10^{-4} . Mini-batches follow $P \times K$ sampling with $P = 8$, $K = 4$ (batch size 32), optimizing batch-hard triplet loss with a margin warm-up $\alpha : 0.2 \rightarrow 0.5$ over the first 40 epochs (total 50 epochs). All reproduced baselines (when applicable) are trained under the same optimizer, learning-rate schedule, number of epochs, and batch size. Hyper-parameters are selected using the validation split only, and the test split is used exactly once for evaluation.

Data augmentation: To improve robustness, we apply small in-plane rotation ($\pm 5^\circ$), brightness/contrast jitter ($\pm 20\%$), light Gaussian blur, random horizontal occlusion bands (to emulate eyelids), and localization jitter (small perturbations of estimated centers/radii) to simulate detection noise.

4.4 Results and Discussions

Table 2 summarizes the intra-dataset verification and identification results across three CASIA-V4 subsets. RadialFormer consistently achieves low EER and high TPR at stringent operating points, demonstrating robust discrimination under both controlled (Interval) and challenging illumination conditions (Lamp). The performance variation across datasets reflects the inherent difficulty differences: CASIA-V4-Interval contains lower-resolution images with more controlled acquisition, while CASIA-V4-Lamp exhibits strong specular highlights and illumination changes that challenge traditional methods. Notably, on CASIA-V4-Lamp—the most challenging subset due to its lamp on/off illumination protocol and larger subject population—the proposed method achieves a remarkably low EER of 0.48% and Rank-1 accuracy exceeding 99%, demonstrating the effectiveness of geometry-aware polar-domain processing for handling illumination-induced appearance variations.

Table 2: Intra-dataset verification and identification performance of RadialFormer on CASIA-V4 benchmarks (mean \pm std over three seeds). EER: Equal Error Rate; TPR: True Positive Rate at specified FPR; R1: Rank-1 identification accuracy.

Dataset	EER% \downarrow	TPR@0.1% \uparrow	TPR@1% \uparrow	R1% \uparrow
CASIA-V4-Interval	2.92 \pm 0.10	84.70 \pm 1.20	93.63 \pm 0.45	97.80 \pm 0.30
CASIA-V4-Lamp	0.48 \pm 0.05	98.28 \pm 0.20	99.04 \pm 0.06	99.20 \pm 0.20
CASIA-V4-Thousand	1.50 \pm 0.15	88.50 \pm 1.00	95.80 \pm 0.40	98.65 \pm 0.20

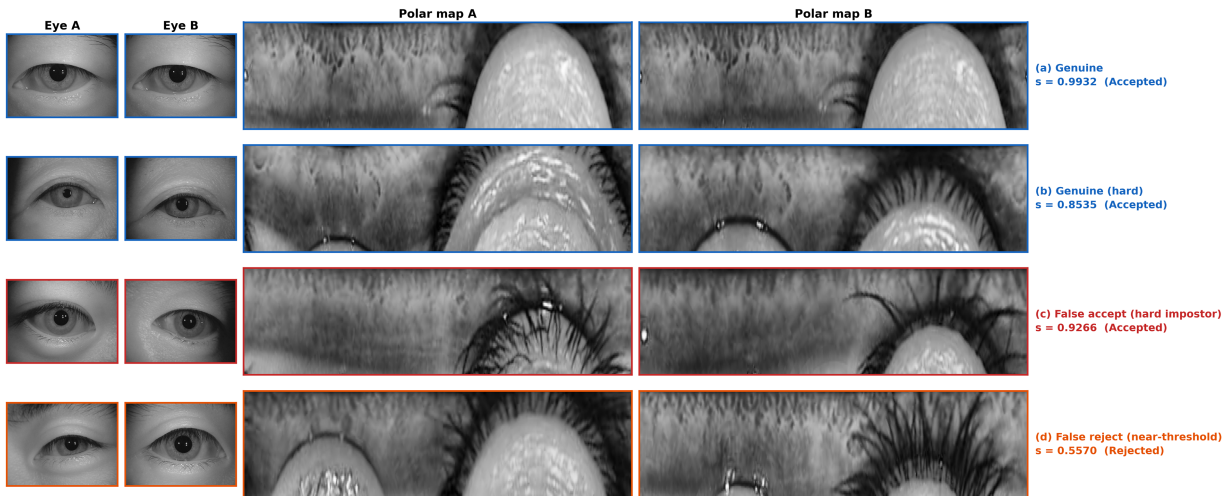
To provide broader context, Table 3 summarizes representative prior results reported on CASIA-IrisV4-Lamp. Because these studies use different protocols and reporting conventions, the table is intended as a contextual comparison rather than a strictly protocol-matched benchmark.

To complement the quantitative results, Fig. 5 presents representative qualitative recognition examples on CASIA-V4-Lamp, including correctly accepted genuine pairs, a hard impostor false-accept case, and a representative false-reject example near the decision threshold. These examples provide a more intuitive view of both the strengths and the practical failure modes of the proposed framework.

Table 3 also clarifies the practical implication of the proposed mask-free design. Representative prior methods on CASIA-IrisV4-Lamp typically rely on segmentation-dependent preprocessing, whereas RadialFormer operates without such preprocessing and still achieves strong verification accuracy. In particular, RadialFormer achieves the best reported performance with 0.48% EER, 98.28% TPR@0.1%FPR, and 99.04% TPR@1%FPR, as highlighted by the bold entries in Table 3. Although these results are contextual rather than strictly protocol-matched, they provide quantitative evidence that competitive recognition performance can be achieved without segmentation-dependent preprocessing.

Table 3: Comparison with representative prior methods on CASIA-IrisV4-Lamp verification, highlighting segmentation dependence and recognition performance. Results for competing methods are reported from their original publications and may not be strictly comparable because of differences in protocols, segmentation settings, and reporting conventions. “—” indicates values not explicitly reported in the corresponding paper.

Method	Segmentation	EER (%) ↓	TPR@0.1% ↑	TPR@1% ↑
IRINA [32]	Required	2.6 ± 0.5	—	—
Yang et al. [33]	Required	2.8 ± 0.4	—	—
Sun and Tan [34]	Required	2.9 ± 0.5	—	—
Belcher and Du [35]	Required	7.7 ± 1.1	—	—
Tahir et al. [36]	Required	2.45	—	96.56
RadialFormer (Ours)	Not required	0.48 ± 0.05	98.28 ± 0.20	99.04 ± 0.06



Decision threshold $\tau = 0.5608$ (at EER = 1.3%). $s \geq \tau \rightarrow$ accepted; $s < \tau \rightarrow$ rejected.

Figure 5: Qualitative recognition examples on CASIA-V4-Lamp. Each row shows a pair of eye images together with their corresponding crop-based polar representations and the cosine similarity score produced by RadialFormer. Rows (a) and (b) illustrate correctly accepted genuine pairs, including a more challenging genuine example. Row (c) shows a representative false-accept case involving a hard impostor pair, while row (d) shows a representative false-reject example near the decision threshold. The decision threshold is set at the EER operating point on the test split.

To evaluate robustness against dataset bias and domain shift, we conduct comprehensive cross-dataset and cross-domain verification experiments across three CASIA-V4 subsets: Lamp, Interval, and Thousand. These subsets differ substantially in acquisition conditions, image quality, and subject distributions, providing a challenging benchmark for assessing generalization. We consider three training strategies to analyze performance under increasing levels of domain mismatch.

In the *single-domain* setting, the model is trained on one subset and directly evaluated on another subset without any fine-tuning. In the *two-domain* setting (denoted as Combined-2), the model is trained on the union of the Lamp and Interval *training splits* and evaluated separately on the test splits of each subset, while the Thousand subset remains unseen during training. In the *three-domain* setting (denoted as Combined-3), the model is trained on the union of the Lamp, Interval, and Thousand *training splits* and evaluated on the corresponding test splits. The Combined-3 setting does not assess unseen-domain generalization; instead, it represents a practical multi-domain deployment configuration and provides an upper-bound performance

estimate when representative data from all target domains are available during training. In all settings, test subjects are strictly disjoint from the training data.

As shown in Table 4, single-domain training exhibits a pronounced performance gap between same-domain and cross-domain evaluation. When trained on Lamp, the model achieves strong performance on Lamp but degrades substantially on Interval, and vice versa. Despite this degradation, RSWA-Single significantly outperforms the CNN baseline in cross-dataset evaluation, improving TPR@1% FPR from 35.17% to 70.11% in the Lamp→Interval setting. These results highlight both the challenge of cross-dataset iris recognition and the improved transferability of geometry-aware polar-domain representations.

Table 4: Single-domain cross-dataset verification performance (TPR@1% FPR). Training and test subjects are disjoint.

Training	Lamp	Interval	Thousand	Gap
Lamp only	99.04%	70.11%	–	28.93%
Interval only	62.89%	93.63%	–	30.74%

Table 5 shows that training on both Lamp and Interval substantially mitigates the domain gap between these two subsets. The Combined-2 model achieves consistently high and balanced accuracy on Lamp and Interval, with TPR@1% FPR exceeding 89% and low EER values below 3% on both domains. When evaluated on the unseen Thousand subset, performance decreases compared to Lamp and Interval, reflecting a significant domain shift caused by lower image quality, severe noise, and different acquisition characteristics. Nevertheless, without any target-domain fine-tuning, the model maintains a TPR of 69.34% at 1% FPR and 45.34% at 0.1% FPR, indicating non-trivial generalization capability under extreme out-of-distribution conditions. This setting therefore serves as a stringent unseen-domain stress test, highlighting both the challenges of cross-domain iris recognition and the robustness of geometry-aware polar-domain representations.

Table 5: Cross-dataset verification performance under two-domain combined training (Combined-2) and three-domain deployment setting (Combined-3).

Test Dataset	Combined-2			Combined-3		
	EER ↓	TPR@0.1% ↑	TPR@1% ↑	EER ↓	TPR@0.1% ↑	TPR@1% ↑
Lamp	2.27%	76.08%	95.33%	2.62%	73.64%	94.09%
Interval	2.83%	–	89.78%	3.59%	–	85.17%
Thousand (zero-shot)	11.63%	45.34%	69.34%	5.40%	61.20%	86.93%

Incorporating all three datasets during training significantly improves robustness, particularly on the challenging Thousand subset. Compared to the zero-shot evaluation in Table 5, including Thousand during training reduces the EER from 11.63% to 5.40% and improves TPR@1% FPR by more than 22 percentage points. These results indicate that a substantial portion of the performance drop observed in the unseen-domain setting is attributable to domain mismatch rather than insufficient model capacity. Notably, the slight performance trade-off on Lamp and Interval compared to single-domain training reflects a typical regularization effect when optimizing for multiple domains simultaneously, rather than a degradation of within-domain discriminability. Importantly, this improvement does not diminish the value of the zero-shot evaluation; instead, it confirms that while domain mismatch accounts for a large fraction of the performance gap, the remaining robustness observed in the unseen-domain setting can be attributed to the geometry-aware polar-domain representation learned by RadialFormer.

Table 6 presents a systematic ablation study on CASIA-V4-Lamp, progressively adding each proposed component to quantify its individual contribution. Switching from Cartesian to polar representation provides a 2.7% absolute improvement in TPR@1%, reducing EER from 4.80% to 3.50%. This gain validates the fundamental premise that polar coordinates better align with iris geometry, even before introducing specialized attention mechanisms. The improvement primarily stems from rotation invariance: in-plane eye rotations become horizontal shifts in the polar domain, which the network can learn to handle through translation-equivariant convolutions.

Table 6: Ablation on CASIA-V4-Lamp (mean over three seeds). Starting from a CNN baseline, we progressively add polar transform, standard positional encoding, LPPE, and RSWA.

Configuration	TPR@1% \uparrow	TPR@0.1% \uparrow	EER (%) \downarrow	Δ TPR@1%
Baseline (CNN only)	88.50	45.20	4.80	–
+ Polar Transform	91.20	58.60	3.50	+2.7
+ Standard Pos. Encoding	94.80	72.30	2.10	+3.6
+ LPPE (Ours)	97.50	89.40	1.20	+2.7
+ RSWA (Ours)	99.04	98.28	0.48	+1.5

Adding standard sinusoidal positional encoding yields an additional 3.6% gain, bringing TPR@1% to 94.80%. This relatively large improvement indicates that explicit position information is critical for discriminating iris textures, as different radial positions carry distinct semantic meaning (pupil boundary features differ from limbus features). Besides, replacing standard positional encoding with our Learnable Polar Position Encoding (LPPE) provides a further 2.7% improvement. The key difference is LPPE’s separable radial-angular decomposition with Fourier components on the angular branch. These Fourier components explicitly enforce wrap-around continuity at $\theta = 0/2\pi$, ensuring that features near angular boundaries receive consistent positional signals. The improvement demonstrates that topology-aware encoding outperforms generic positional representations.

The final addition of Radial Stripe Window Attention (RSWA) yields the complete RadialFormer, achieving 99.04% TPR@1% with 0.48% EER. RSWA provides the largest gain at the stringent TPR@0.1% operating point (from 89.40% to 98.28%), indicating its effectiveness for high-precision verification scenarios. The improvement validates our hypothesis that full-height radial stripes better capture pupil-to-limbus dependencies than square windows. Importantly, neither polar transformation nor LPPE or RSWA alone achieves the full performance gain. The combination of geometry-aware positional encoding and radial-aligned attention is essential for robust iris representation learning. As highlighted by the bold entries in Table 6, the complete RadialFormer configuration achieves the best overall performance with 99.04% TPR@1%FPR, 98.28% TPR@0.1%FPR, and 0.48% EER, confirming that both LPPE and RSWA contribute meaningful improvements to discriminative iris representation learning.

Fig. 6 evaluates the robustness of RadialFormer under synthetic perturbations applied before polar unwrapping. As shown in Fig. 6a, the model maintains consistently high TPR@1% FPR across a wide range of in-plane rotations ($\pm 25^\circ$), confirming the rotation-consistent design of LPPE and RSWA. The performance remains above 97% TPR even at extreme rotation angles of $\pm 25^\circ$, demonstrating that the polar transformation successfully converts rotations into horizontal shifts that RSWA can handle through its modular wrap-around mechanism.

Fig. 6b shows that performance degrades gracefully as the occlusion ratio increases, demonstrating resilience to eyelid-like occlusions and partial iris corruption. At 20% horizontal occlusion—simulating

typical eyelid coverage—TPR@1% remains above 95%, validating the robustness of RSWA’s stripe-based attention which can leverage unoccluded radial stripes for matching.

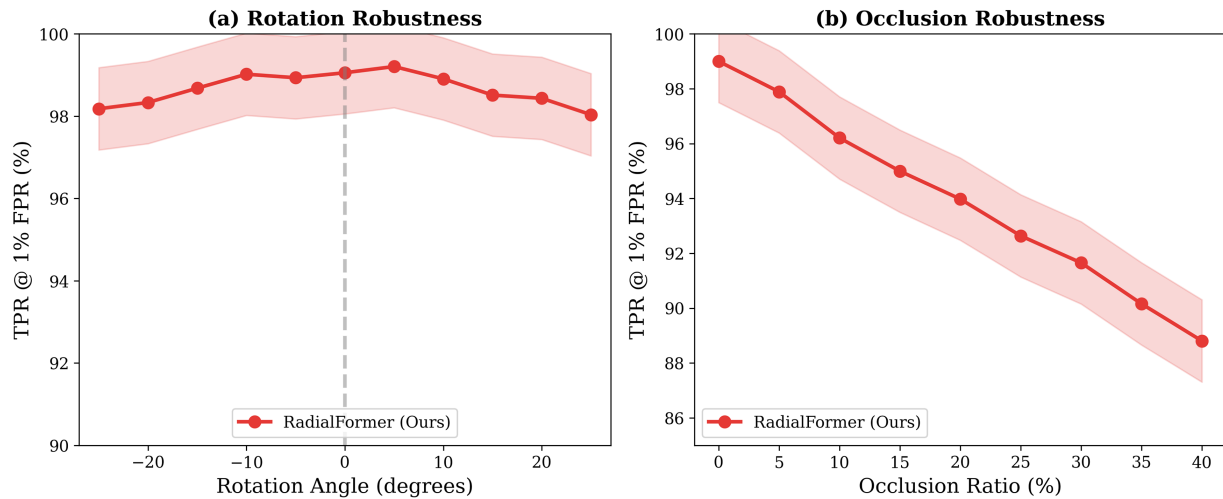


Figure 6: Robustness on CASIA-V4-Lamp: (a) rotation robustness using synthetic in-plane rotations prior to unwrapping; (b) occlusion robustness under increasing horizontal occlusion ratios. Performance is evaluated as TPR at 1% FPR.

Table 7 compares model complexity and inference throughput. RadialFormer achieves $3.5\times$ fewer FLOPs than ViT-B/16 and $2.4\times$ fewer than Swin-Tiny, while using only 7.8M parameters compared to 28.3M and 86.6M for the baselines. This efficiency stems from three factors: (i) crop-based polar transformation reduces the effective input resolution by eliminating background regions; (ii) asymmetric downsampling along the radial axis reduces token count while preserving discriminative angular information; and (iii) RSWA’s stripe-based partitioning enables linear-complexity attention with favorable constants. At 380 FPS on a single GPU, RadialFormer is suitable for real-time applications in border control and mobile authentication. The bold entries in **Table 7** indicate that RadialFormer achieves the most favorable balance between model size, computational cost, and throughput, requiring only 7.8M parameters and 1.9 GFLOPs while reaching 380 FPS.

Table 7: Model complexity and inference throughput on a single NVIDIA RTX 3060 GPU under the same input configuration.

Model	Params (M)	GFLOPs	FPS
ResNet-50 [37]	25.6	4.1	285
Swin-Tiny [11]	28.3	4.5	245
ViT-B/16 [10]	86.6	17.6	125
RadialFormer (Ours)	7.8	1.9	380

Fig. 7 presents the ROC curve on CASIA-V4-Lamp. The steep rise at low false positive rates indicates strong discriminative power, consistent with the low EER and high TPR evaluated in **Table 2**. The curve approaches the upper-left corner rapidly, demonstrating that RadialFormer produces well-separated embedding distributions for genuine and impostor pairs.

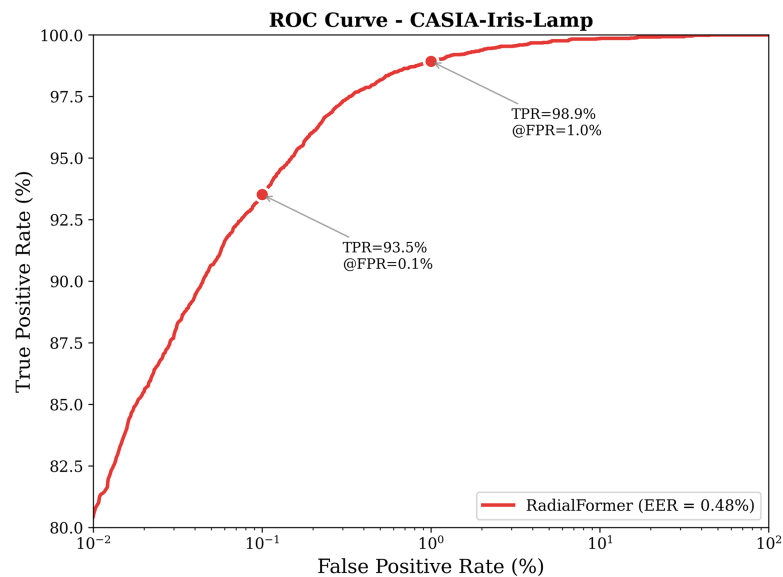


Figure 7: ROC curve on CASIA-V4-Lamp with operating points at FPR = 0.1% and FPR = 1%.

Beyond the quantitative gains, the present results suggest several practical implications of the proposed design. First, explicitly aligning representation learning with iris geometry yields a favorable balance between accuracy and efficiency. The strong performance on CASIA-V4-Lamp and CASIA-V4-Interval, together with the ablation gains of polar transformation, LPPE, and RSWA, indicates that geometry-aware positional encoding and stripe-aligned attention are effective for contactless iris verification under challenging illumination and appearance variation.

At the same time, the proposed framework retains several limitations. Although it does not require pixel-wise segmentation masks, it still depends on the quality of the mask-free localization stage. When center or radius estimation is inaccurate, the resulting polar representation may be degraded, which can lead to borderline false rejects or unstable similarity scores, as also illustrated by the qualitative examples. In addition, the current crop-based unwrapping uses a shared-center approximation for pupil and limbus boundaries. While this choice improves simplicity and efficiency and works well under the predominantly near-frontal NIR acquisition conditions of CASIA-IrisV4, it may be less suitable for more strongly off-axis imagery or cases with pronounced pupil–limbus decentering.

The cross-dataset results further show that domain mismatch remains a meaningful challenge. Although the proposed representation generalizes reasonably well, especially under Combined-2 and Combined-3 training, unseen-domain performance is still lower than within-domain performance. This suggests that geometry-aware design improves robustness but does not eliminate the need for broader domain coverage. Future work will therefore investigate more flexible non-concentric normalization, lightweight learned localization modules, and evaluation on more diverse cross-sensor and off-angle iris benchmarks.

5 Conclusion

This paper proposed RadialFormer, a segmentation-free iris recognition framework that performs representation learning directly in the polar domain and explicitly models the circular geometry of iris patterns. The key motivation is that, although classical iris recognition has long used polar normalization, many modern deep encoders still inherit assumptions from generic Cartesian image modeling: iris texture is organized concentrically from pupil to limbus and exhibits angular periodicity, whereas many deep encoders

assume rectangular, non-periodic spatial structure. This mismatch can make rotation handling less efficient and can weaken pupil-to-limbus context modeling, particularly when segmentation or boundary estimation is unreliable.

RadialFormer addresses these issues with two geometry-aware encoder components. Learnable Polar Position Encoding (LPPE) decomposes position into separable radial and angular embeddings, and augments the angular branch with Fourier terms to encode wrap-around continuity at $\theta = 0/2\pi$. Radial Stripe Window Attention (RSWA) performs self-attention within full-height radial stripes to capture complete radial dependencies in each window, and uses modular angular shifting to preserve circular consistency while enabling cross-stripe interaction.

On three CASIA-IrisV4 NIR benchmarks, RadialFormer achieves competitive verification and identification performance without relying on pixel-wise segmentation masks. In particular, on CASIA-V4-Lamp it reaches 99.04% TPR@1%FPR with 0.48% EER. Under the same input resolution, the proposed design reduces computation by about 3.5× compared with a transformer baseline while maintaining strong recognition accuracy. Ablation studies further verify that polar processing, LPPE, and RSWA each contribute measurable improvements.

The proposed mask-free localization relies on analytic cues and may degrade under extreme off-angle gaze, severe blur, or heavy occlusions. In addition, the evaluation focuses on NIR imagery within CASIA-V4; broader validation on more diverse benchmarks and sensors is a natural next step. Future work will explore a lightweight differentiable center/radius regressor to enable closer-to-end-to-end optimization and improved robustness.

Acknowledgement: This research is supported by the Posts and Telecommunications Institute of Technology (PTIT), Vietnam.

Funding Statement: Not applicable.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Trong-Thua Huynh and De-Thu Huynh; methodology, Trong-Thua Huynh and Cong-Sang Duong; software: Cong-Sang Duong and De-Thu Huynh; validation, Quoc H. Nguyen, Lam-Thanh Tu and Hong-Son Nguyen; formal analysis, Trong-Thua Huynh and Quoc H. Nguyen; resources, Cong-Sang Duong; data curation, Trong-Thua Huynh; writing—original draft preparation, Cong-Sang Duong and De-Thu Huynh; writing—review and editing, Trong-Thua Huynh and Cong-Sang Duong; visualization, Lam-Thanh Tu; supervision, Hong-Son Nguyen; project administration, Trong-Thua Huynh. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used in this study are from the Institute of Automation of the Chinese Academy of Sciences (CASIA), http://english.ia.cas.cn/db/201610/t20161026_169399.html. Researchers who wish to obtain the original dataset should email the official provider directly. The source code and trained models are available at https://drive.google.com/drive/folders/1PsgJIsRmmc-wKo9OXhRPBp8ovZEznJEJ?usp=drive_link. To preserve the integrity of the peer-review process, the repository is currently private; however, researchers seeking the source code to reproduce our results may request access from the corresponding author, who will grant permission upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Daugman J. How Iris recognition works. *IEEE Trans Circ Syst Video Technol.* 2004;14(1):21–30. doi:10.1109/TCSVT.2003.818350.
2. Bowyer KW, Hollingsworth K, Flynn PJ. *Handbook of Iris recognition*. 2nd ed. Cham, Switzerland: Springer; 2016.
3. Daugman JG. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans Pattern Anal Mach Intell.* 1993;15(11):1148–61. doi:10.1109/34.244676.
4. Proença H, Alexandre LA. Iris recognition: on the segmentation of degraded images acquired in the visible wavelength. *IEEE Trans Pattern Anal Mach Intell.* 2010;32(8):1502–16.
5. Wildes RP. Iris recognition: an emerging biometric technology. *Proc IEEE.* 1997;85(9):1348–63. doi:10.1109/5.628669.
6. Arsalan M, Hong HG, Naqvi RA, Lee MB, Kim MC, Kim DS, et al. Deep learning-based iris segmentation for iris recognition in visible light environment. *Symmetry.* 2017;9(11):263. doi:10.3390/sym9110263.
7. Lozej J, Meden B, Struc V, Peer P. End-to-end Iris segmentation using U-Net. In: *Proceedings of the IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*. Piscataway, NJ, USA: IEEE; 2018. p. 1–6.
8. Gangwar A, Joshi A. DeepIrisNet: deep Iris representation with applications in Iris recognition and cross-sensor Iris recognition. In: *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. Piscataway, NJ, USA: IEEE; 2016. p. 2301–5.
9. Nguyen K, Fookes C, Jillela R, Sridharan S, Ross A. Long range Iris recognition: a survey. *Pattern Recognit.* 2017;72:123–43. doi:10.1016/j.patcog.2017.05.021.
10. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 words: transformers for image recognition at scale. *arXiv:2010.11929*. 2020.
11. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway, NJ, USA: IEEE; 2021. p. 10012–22.
12. Othman N, Dorizzi B, Garcia-Salicetti S. OSIRIS: an open source Iris recognition software. *Pattern Recognit Lett.* 2016;82:124–31. doi:10.1016/j.patrec.2015.09.002.
13. Shah S, Ross A. Iris segmentation using geodesic active contours. *IEEE Trans Inf Forensics Secur.* 2009;4(4):824–36. doi:10.1109/tifs.2009.2033225.
14. Roy K, Bhattacharya P, Suen CY. Iris segmentation using variational level set method. *Opt Lasers Eng.* 2011;49(4):578–88. doi:10.1016/j.optlaseng.2010.09.011.
15. Daugman J. New methods in Iris recognition. *IEEE Trans Syst Man Cybern B Cybern.* 2007;37(5):1167–75. doi:10.1109/tsmcb.2007.903540.
16. 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;15:2944–59. doi:10.1109/tifs.2020.2980791.
17. Lakshmi S, Sankaranarayanan V, Hanumanthappa M. IrisDenseNet: robust Iris segmentation using densely connected fully convolutional networks. *Expert Syst Appl.* 2018;112:68–79.
18. Zhao T, Liu Y, Huo G, Zhu X. A deep learning Iris recognition method based on capsule network architecture. *IEEE Access.* 2019;7:49691–701. doi:10.1109/ACCESS.2019.2911056.
19. Gangwar A, Joshi A. DeepIrisNet2: learning deep Iris representations for Iris recognition from a large-scale dataset. *arXiv:1907.09380*. 2019.
20. Lei S, Dong B, Shan A, Li Y, Zhang W, Xiao F. Attention meta-transfer learning approach for few-shot Iris recognition. *Comput Electr Eng.* 2022;99:107848. doi:10.1016/j.compeleceng.2022.107848.
21. Wei Z, Tan T, Sun Z. Towards more discriminative and robust Iris recognition by learning uncertain factors. *IEEE Trans Inf Forensics Secur.* 2022;17:865–79.
22. Alinia Lat R, Danishvar S, Heravi H, Danishvar M. Boosting Iris recognition by margin-based loss functions. *Algorithms.* 2022;15(4):118. doi:10.3390/a15040118.

23. Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2015. p. 815–23.
24. Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv:1703.07737. 2017.
25. Song HO, Xiang Y, Jegelka S, Savarese S. Deep metric learning via lifted structured feature embedding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2016. p. 4004–12.
26. Sohn K. Improved deep metric learning with multi-class N-pair loss objective. In: Advances in neural information processing systems (NeurIPS). Red Hook, NY, USA: Curran Associates, Inc.; 2016. p. 1857–65.
27. Wang X, Han X, Huang W, Dong D, Scott MR. Multi-similarity loss with general pair weighting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2019. p. 5022–30.
28. Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2019. p. 4690–9.
29. Zhao Z, Kumar A. Towards more accurate Iris recognition using deeply learned spatially corresponding features. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE; 2017. p. 3752–60. doi:10.1109/ICCV.2017.411.
30. Zhang Y, Zhou Z, David P, Yue X, Xi B, Gong B, et al. PolarNet: an improved grid representation for online LiDAR point clouds semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2020. p. 9601–10.
31. Jiang Y, Zhang L, Miao Z, Zhu X, Gao J, Hu W, et al. PolarFormer: multi-camera 3D object detection with polar transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2023. p. 1042–51.
32. Proença H, Neves JC. IRINA: Iris recognition (even) in inaccurately segmented data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2017. p. 6700–9.
33. Yang G, Zeng H, Li P, Zhang L. High-order information for robust Iris recognition under less controlled conditions. In: Proceedings of the IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE; 2015. p. 4535–9.
34. Sun Z, Tan T. Ordinal measures for Iris recognition. IEEE Trans Pattern Anal Mach Intell. 2009;31(12):2211–26. doi:10.1109/tpami.2008.240.
35. Belcher C, Du Y. Region-based SIFT approach to Iris recognition. Opt Lasers Eng. 2009;47(1):139–47.
36. Tahir AAK, Dawood S, Anghelus S. An Iris recognition system using a new method of Iris localization. Int J Open Inf Technol. 2021;9(6):41–9.
37. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2016. p. 770–8. doi:10.1109/CVPR.2016.90.