arXiv:2608.08977v1 [cs.CV] 10 Aug 2026

Paper accepted at IEEE/IAPR International Joint Conference on Biometrics (IJCB), September 2026.

Detecting Clear Contact Lenses for Iris Recognition: A Two-Stage Mask-Guided Attention Approach

Parisa Farmanifard and Arun Ross
Michigan State University, East Lansing, MI 48824
{farmanif,rossarun}@msu.edu

Abstract

This work focuses on the impact and detection of clear contact lenses in the context of iris recognition. While the detection of cosmetic or patterned contact lenses has been extensively studied under the presentation attack detection (PAD) paradigm, clear prescription contact lenses, that are typically transparent, have received comparatively less attention despite their widespread use. Unlike patterned lenses, clear lenses introduce no salient texture artifact, making them difficult to detect and are often assumed to have no impact on iris recognition. We first examine this assumption using the commercial VeriEye matcher on four benchmark datasets and show that clear lenses marginally degrade genuine match scores and increase verification error. We then propose a two-stage contact-lens detection framework. Stage 1 uses an existing PAD model to identify patterned lenses, while Stage 2 focuses on the more challenging clear-lens versus no-lens distinction using a ConvNeXt-Base model equipped with Mask-Guided Spatial Attention (MGSA). The proposed MGSA module incorporates a Hough-derived anatomical ROI mask together with learned spatial attention and Squeeze-and-Excitation channel recalibration, allowing the network to focus on subtle limbal cues associated with clear lens wear. Across four datasets, the full pipeline consisting of both patterned and clear contact lens detection achieves between 90.0%–98.8% accuracy. Finally, we introduce a z-score calibration method that adjusts VeriEye match scores when a clear lens is detected in the input images. This calibration reduces EER by 4.1%–28.3% across datasets, demonstrating that reliable clear contact lens detection can directly improve iris verification performance.

1. Introduction

The iris is the annular region of the eye surrounding the pupil and is a powerful biometric cue whose intricate texture can even be used to distinguish identical twins [1, 2]. However, this efficacy depends on an implicit assumption: that the captured iris is an unobstructed, unaltered view of bona fide ocular tissue. Contact lenses–including patterned cosmetic lenses and the far more prevalent clear[1] lenses–challenge this assumption in clinically distinct ways (samples can be seen in Fig.1). This is a widespread concern, as more than 140 million people worldwide were estimated to wear contact lenses in 2025 [4]. Patterned lenses are a well-studied threat. Cosmetic (patterned) contact lenses *typically* overlay a bold[2] synthetic texture over the iris, cor-

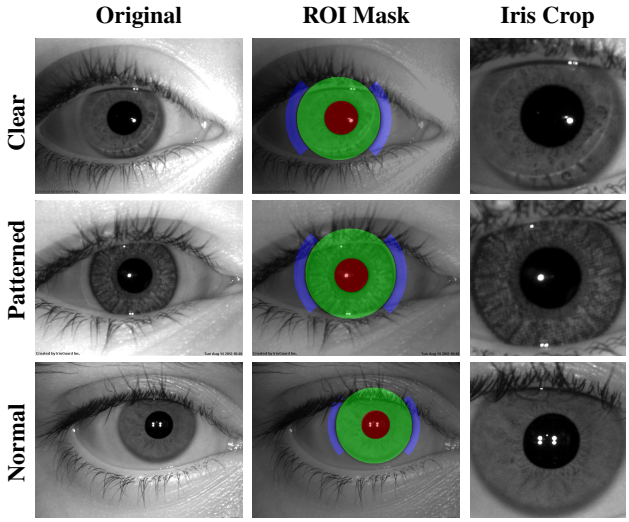

Figure 1. Sample inputs to the proposed pipeline for each lens class. Each row shows the full periocular image, the anatomical ROI mask M (Eq. 3), and the localised iris crop $\hat{I}$ used by Stage 1. In the mask column: **red** = pupil region, **green** = iris annulus, **blue** = outer arc beyond the limbus where clear-lens evidence appears.

[1] In the iris presentation attack detection (PAD) literature, clear contact lenses are also commonly referred to as transparent or soft contact lenses [3].

[2] Although most cosmetic contact lenses studied in the iris biometrics literature have a "strong" textured pattern, there are other such lenses with subtle texture patterns.

rupting the Gabor-phase feature codes used for iris comparison and constituting a well-documented presentation attack [5, 6, 7, 8, 9, 10]. They are detectable precisely because the foreign texture is often visually conspicuous: a compact, well-aligned iris crop reveals the periodic pattern, and even lightweight classifiers can learn to separate it from a natural iris with high reliability [11, 6, 12, 13]. On the other hand, clear lenses are an underexplored challenge. Clear contact lenses–worn by people for vision correction–have attracted far less attention in the iris literature, on the assumption that an optically clear lens leaves the iris appearance unchanged. This assumption does not always hold in practice. Baker et al. [14] first showed, across more than 12,000 images, that clear lenses produce false non-matches through subtle deformation and edge artifacts at the iris limbus. Later work also confirmed that this degradation is not random noise but a systematic, condition-dependent bias: the refractive index mismatch at the lens boundary creates a predictable reflectance deviation in the periocular ring [15, 16]. Later, another paper [17], proposed ContlensNet, a method to detect clear contact lenses in ocular images.

Despite over a decade of research, detecting clear contact lenses still remains an open problem [18]. The difficulty arises from a combination of factors that together make it a challenging fine-grained classification task in biometrics:

- Optical invisibility: Clear lenses carry no colour or texture signature, unlike cosmetic contact lenses. The only physical evidence is a faint reflectance deviation at the iris–sclera limbus, invisible to the naked eye and easily confused with natural periocular variation.
- Extreme spatial localization: Clear lens evidence is often confined to a narrow annular band at the lens edge. Standard classification backbones trained to aggregate global information tend to ignore this small region in favour of more salient but uninformative background features.
- High intra-class variability: Different lens brands, geometries, and moisture levels produce different limbal signatures, further modulated by eye colour, pupil dilation, and illumination angle.
- Two fundamentally different sub-problems under one label: Patterned and clear lenses share very less commonality. A detection model trained on the bold periodic textures of a cosmetic lens is suboptimal for clear lenses.
- Sensor and acquisition variability: Iris sensors differ in resolution, near-infrared wavelength, and depth of field, all of which alter the appearance of the lens edge and surrounding periocular tissue across devices.

Our approach, motivated by the fundamental disparity between the two detection sub-problems, i.e., cosmetic patterned lens detection and clear lens detection, utilizes a two-stage method pipeline. Stage 1 uses an existing iris PAD model known as D-NetPAD [19] to screen for patterned lens artefacts from a compact iris crop and passing only non-patterned images to Stage 2. Stage 2 introduces a novel Mask-Guided Attention Network that injects an anatomical ROI mask derived from classical Hough-based iris localization [2], with no learned segmentation and no pixel-level annotation, directly into the backbone's multi-level feature representation. The mask encodes the hard spatial prior that all clear lens evidence is often confined to the iris annulus and the periocular arc at the limbal boundary. Fusing this prior multiplicatively into the feature representation forces the network to commit to the anatomically correct region, eliminating the shortcut of latching onto uninformative background cues. A learned spatial gate within the same module further modulates which parts of the masked region are most diagnostic, while a Squeeze-and-Excitation block [20] recalibrates feature channels globally–together asking *where* to look and *what* to look for. We implement Stage 2 with three backbone architectures to assess how discriminative capacity and receptive field size interact with mask guidance: EfficientNet-B4 [21], ResNet-101 [22], and ConvNeXt-Base [23]. Once a clear contact lens is detected, we apply a z-score calibration for adjusting verification scores.

This work makes the following contributions:

1. Quantitative evidence that clear lenses degrade iris verification. We quantify, across four large-scale datasets using a commercial iris-matching SDK (VeriEye) [24], the systematic genuine-score drop introduced by clear lenses.
2. A two-stage method for contact-lens detection. Decomposes the three-class classification problem ("normal", "clear", "patterned") into a texture sub-problem (Stage 1, patterned vs. non-patterned) and a spatial localization sub-problem (Stage 2, clear vs. normal), substantially outperforming single-stage baselines.
3. A novel mask-guided attention mechanism. Injects an annotation-free anatomical ROI mask into the intermediate feature maps of a deep backbone. This combines a learned spatial attention gate with a fixed anatomy-based prior, allowing the contact-lens detector to focus on the most relevant iris regions during feature learning. Moreover, a backbone study comparing EfficientNet-B4, ResNet-101, and ConvNeXt-Base under identical mask-guidance conditions, showing that large-kernel depthwise convolutions best resolve the subtle limbal evidence of clear lens wear.
4. Downstream verification improvement. Upto 28.3% relative EER reduction via predicted-label score calibration.

2. Related Work

Iris Presentation Attack Detection: Iris PAD has been an active area of research since the discovery that printed irides, replayed videos, and textured contact lenses can successfully fool commercial matchers [25, 26]. Early methods extracted hand-crafted descriptors, viz., BSIF, LBP, and co-occurrence matrices, from the iris annulus and fed

them to shallow classifiers [27]. Deep learning substantially raised the bar: Menotti et al. [11] showed that convolutional networks fine-tuned end-to-end on iris crops outperform handcrafted pipelines across multiple attack types. D-NetPAD [19] built on DenseNet-121 [28] with interpretability constraints, demonstrated strong generalization to cosmetic lens artifacts across sensors [29]; we adopt it as our Stage 1 patterned-lens filter. More recent work has explored multi-task and attention-based formulations. Chen and Ross [30] proposed an attention-guided PAD framework that localizes discriminative iris regions using channel and spatial attention gates. The LivDet-Iris competition series [31, 29, 32] provides standardised cross-sensor benchmarks; the 2023 edition introduced GAN-generated irides as an attack category, highlighting the growing diversity of the threat landscape. The recent top-performing LivDet-Iris competition [33] achieved good performance, showing that PAD is a well-studied problem for patterned lenses.

Clear Lens Detection: Unlike patterned contact lenses, clear lenses leave no strong texture signature, making detection fundamentally harder. Baker et al. [14] established the empirical baseline, demonstrating that clear lenses increase false non-match rates even for high-quality NIR sensors. Doyle and Bowyer [15] later showed that the lens boundary introduces a measurable 3D surface distortion detectable with photometric stereo, though this approach requires controlled multi-light acquisition. Raghavendra et al. [17] proposed ContlensNet, a 15-layer CNN that classifies normalized iris patches as no lens, soft lens, or textured lens and determines the image-level label through majority voting. Yadav et al. [16] conducted a large-scale analysis of the effect of textured and incidentally clear lenses on commercial matchers, quantifying the score drop across multiple cohorts. Unlike ContlensNet [17], which only reports lens classification performance, our work goes beyond that by reporting iris recognition and score calibration results. Further, we use the pre-normalized ocular image, which includes the scleral region, thereby capturing the outer boundary of lenses.

Attention in Deep Recognition Models: Attention mechanisms have become a standard tool for directing network capacity to informative spatial regions [20, 34]. The Squeeze-and-Excitation (SE) block [20] introduced channel-wise recalibration via global average pooling, allowing the network to selectively amplify informative feature channels. CBAM [34] extended this with a sequential spatial gate, computing a 2D attention map from both average- and max-pooled features. Non-local networks [35] and Vision Transformers [36] model long-range dependencies via self-attention, but at a cost in data efficiency on small biometric datasets. In the medical imaging domain, mask-guided attention networks have been proposed to incorporate anatomical segmentations as spatial priors, improving sensitivity to lesion regions without requiring lesion-level supervision [37]. Our approach is inspired by this line of work: rather than learning attention maps entirely from the training signal, we inject a Hough-derived iris ROI mask as a hard prior directly into the backbone's intermediate feature hierarchy, coupling it with a learned spatial gate and a SE recalibration block. This is, to our knowledge, the first application of anatomy-driven external mask guidance to contact-lens detection.

Backbone Architectures: ResNet [22] established deep residual learning as the standard baseline for visual recognition, and it remains competitive on biometric tasks due to its simplicity and well-understood inductive bias. EfficientNet [21] introduced compound coefficient scaling of width, depth, and resolution, achieving strong accuracy-efficiency trade-offs on ImageNet while keeping parameter counts low attractive for resource-constrained biometric deployment. ConvNeXt [23] revisited the ResNet design space in light of Vision Transformer innovations, replacing 3×3 convolutions with 7×7 depthwise kernels, adopting LayerNorm, and applying the inverted major challenge structure; the resulting architecture matches or exceeds Swin Transformer on standard benchmarks while remaining fully convolutional. For our task, the enlarged receptive field of ConvNeXt is especially relevant: the limbal evidence of a clear lens spans a narrow but spatially extended annular band, and larger kernels are better positioned to capture the subtle curvature and reflectance gradients that distinguish a lens boundary from natural periocular tissue. We compare all three backbones under identical mask-guidance conditions to isolate the contribution of receptive field size to clear-lens discriminability.

3. Proposed Methodology

3.1. Pipeline Overview

Contact lens detection is not a single problem; it is two fundamentally different problems sharing the same output space. Patterned lenses introduce a strong global texture anomaly detectable from a compact iris crop, while clear lenses leave only a faint annular trace in the vicinity of the limbal boundary. Treating both with a single three-class classifier forces the network to simultaneously learn coarse texture suppression and subtle annular contrast differences, which compete for representational capacity and degrade performance on the harder class. We address this asymmetry with a two-stage method that routes each image through two specialized binary decisions:

$$\hat{y} = \begin{cases} \texttt{patterned}, & \text{if } f_1(\hat{I}) = \texttt{patterned}, \\ \texttt{clear}, & \text{if } f_1(\hat{I}) = \texttt{non-patterned} \\ & \quad \wedge\ f_2(I, M) = \texttt{clear}, \\ \texttt{normal}, & \text{otherwise.} \end{cases} \tag{1}$$

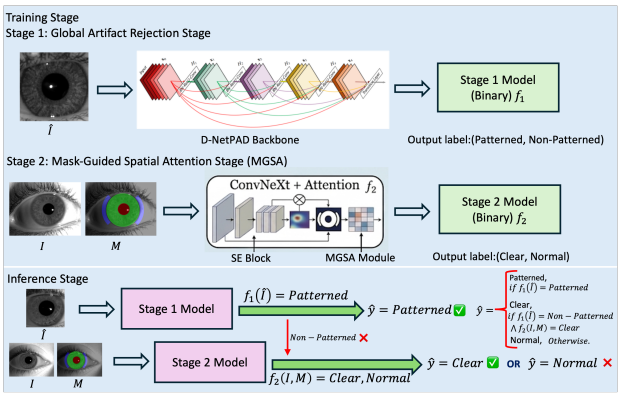

Figure 2. Two-Stage Contact Lens Detection Framework. Top: Stage 1 uses cropped iris images to detect patterned lenses, while Stage 2 uses full periocular images and ROI masks to classify clear vs. normal lenses. Bottom: At inference, images predicted as non-patterned by Stage 1 are passed to Stage 2 for final classification. I denotes the full iris image, M denotes the corresponding mask, and $\hat{I}$ denotes the cropped iris image.

where, $\hat{I}$ is a localized iris crop fed to Stage 1, I is the full periocular image fed to Stage 2, and M is an automatically derived anatomical ROI mask. $f_1 : \hat{I} \rightarrow$ `patterned, non-patterned` is the Stage 1 D-NetPAD binary classifier, and $f_2 : (I, M) \rightarrow$ `clear, normal` is the Stage 2 ConvNeXt-Base + MGSA classifier. Positive Stage 1 detections exit immediately; only non-patterned images reach the computationally heavier Stage 2, keeping the clear-lens sub-problem free from easy cases that would otherwise dominate the gradient. Fig. 2 illustrates the complete pipeline.

3.2. Iris Masks

The fundamental spatial prior for clear-lens detection is that all evidence is confined to the iris and the narrow band immediately surrounding it at the limbus. We encode this prior as a binary mask $M \in \{0,1\}^{H \times W}$ derived entirely from classical iris geometry so that no learned segmentation and no pixel-level annotation are required. For pupil and limbus localization, we apply the circular Hough transform [38] twice to each greyscale image. The circular Hough transform [38, 39] is a voting-based method that finds circles in an edge image by accumulating evidence for all (x, y, r) triples consistent with detected edge points. For the pupil, the image is first median-blurred and thresholded to isolate the dark pupil disk; Canny edges are extracted and fed to `HoughCircles`; the accumulator threshold is lowered iteratively until at least ten candidate circles are returned. The final pupil circle $(c_x^p,\, c_y^p,\, r_p)$–where c_x^p, c_y^p are the horizontal and vertical coordinates of the pupil centre and r_p is its radius–is the coordinate-wise mean of the accepted candidates, which suppresses outliers from eyelash occlusion. The limbus circle $(c_x^\ell,\, c_y^\ell,\, r_\ell)$ is found with the same multi-scale search, constrained so that its centre lies within $0.25\, r_p$ of the pupil centre and its radius exceeds $1.5\, r_p$. Three-region labelling. Given the two detected circles, every pixel $\mathbf{p} = (x, y)$ is assigned one of three anatomical labels:

$$\ell(\mathbf{p}) = \begin{cases} \texttt{pupil}, & \text{if } \|\mathbf{p} - \mathbf{c}^p\|_2 \le r_p, \\ \texttt{iris}, & \text{if } r_p < \|\mathbf{p} - \mathbf{c}^\ell\|_2 \le r_\ell, \\ \texttt{outer_arc}, & \text{if } r_\ell < \|\mathbf{p} - \mathbf{c}^\ell\|_2 \le r_\ell + \delta \\ & \quad \wedge\ \theta(\mathbf{p}) \in \Theta, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The `outer_arc` region captures the periocular tissue immediately beyond the limbus, where the lens edge creates its most visible reflectance deviation. It is restricted to the inferior and superior quadrants $\Theta = [45°, 135°] \cup [225°, 315°]$ (measured from the horizontal), since eyelid occlusion dominates the nasal and temporal sectors and provides no lens cue. The radial offset $\delta = 35$ px beyond the detected limbus accommodates localization uncertainty. The final binary mask is the union of all three labelled regions:

$$M(\mathbf{p}) = \mathbf{1}[\ell(\mathbf{p}) \in \{\texttt{pupil}, \texttt{iris}, \texttt{outer_arc}\}]. \tag{3}$$

Fig. 1 shows representative masks for each lens class. However, the generated masks are not always robust, leaving room for further improvement. Since mask quality directly affects the final classification decision, accurate mask generation is crucial to the overall performance of the proposed method.

3.3. Stage 1 | Patterned Lens Screening

Stage 1 solves a binary classification problem: does the image contain a patterned (cosmetic) lens? Patterned lenses impose a bold, periodic synthetic texture over the iris that is immediately apparent in a well-aligned iris crop $\hat{I}$ of size 256×256. We adopt D-NetPAD [19], a DenseNet 121 [28] fine-tuned specifically for iris presentation attack detection, as our Stage 1 classifier, f_1. D-NetPAD is pretrained on a large pool of iris PAD data and generalizes well to the cosmetic lens texture anomaly; we fine-tune only its classification head on each target dataset. Images predicted as patterned exit the pipeline immediately with label `patterned`. Images not predicted as patterned are passed to Stage 2, which now operates on a cleaner distribution free of the dominant texture cue.

3.4. Stage 2 | Mask-Guided Attention Network

Stage 2 addresses the truly challenging sub-problem: distinguishing a bare iris (`normal`) from one wearing a clear lens (`clear`). Unlike patterned lenses, which alter the global iris texture, clear lenses preserve it for the most part. The only physical evidence of their presence is a faint reflectance discontinuity confined to the narrow annular band at the iris limbus. Two design decisions follow

directly from this observation. First, we use the *full* periocular image I at its native resolution (480×640) rather than a tight iris crop, to preserve the spatial extent of the limbal region. Second, we introduce a Mask-Guided Spatial Attention (MGSA) module that injects the anatomical mask M directly into the backbone's intermediate feature hierarchy, preventing the model from attending to periocular background structures that carry no lens signal.

3.4.1 Backbone

We instantiate Stage 2 with ConvNeXt-Base [23] pretrained on ImageNet-1K [40]. The ConvNeXt architecture consists of a stem followed by four stages with progressively downsampled feature maps; channel dimensions for the Base variant are $[128, 256, 512, 1024]$ at stages 0–3 respectively. We inject MGSA after Stage 2 of the backbone (C=512 channels), where the feature map retains sufficient spatial resolution to resolve the limbal cues while encoding semantically meaningful mid-level representations. Following the attention module, a global average pool and a dropout layer (p=0.4) precede the two-class linear head. We compare ConvNeXt-Base against ResNet-101 [22] and EfficientNet-B4 [21] under identical MGSA conditions in our ablation study (Sec. 4).

3.4.2 Mask-Guided Spatial Attention (MGSA)

Let $\boldsymbol{F} \in \mathbb{R}^{B\times C\times H'\times W'}$ be the feature map produced after the chosen backbone stage. The core insight driving MGSA is that two complementary signals are needed to localize the lens cue: (i) *where* to look, encoded by the anatomical mask M, and (ii) *what to look for*, encoded by a learned spatial saliency gate. Neither signal alone is sufficient: the mask without learning provides no discriminative weighting within the iris, and a learned gate without the mask may attend to eyelid or periocular skin features that correlate with the label only due to dataset biases. The mask is first binarized and upsampled by nearest-neighbour interpolation to match the feature-map resolution:

$$\tilde{M} = \mathrm{Upsample}(M,\ (H', W')) \ \in\ \{0,1\}^{H'\times W'}. \quad (4)$$

A learned saliency gate is produced by a single 1×1 convolution that collapses the channel dimension to a scalar confidence map:

$$G = \sigma\big(\mathbf{w}_g * \boldsymbol{F}\big) \ \in\ (0,1)^{1\times H'\times W'}, \quad \mathbf{w}_g \in \mathbb{R}^{1\times C\times 1\times 1}. \quad (5)$$

The MGSA output is the element-wise triple product:

$$\boxed{\boldsymbol{F}^{*} = \boldsymbol{F} \ \odot\ G \ \odot\ \tilde{M},} \quad (6)$$

where $\odot$ denotes spatially broadcast element-wise multiplication. $\tilde{M}$ acts as a *hard* anatomical gate: it unconditionally zeros all activations outside the iris region, regardless of what the network has learned. G provides a *soft* learned reweighting of positions within that constrained region, learning to further up-weight the outer annulus and limbal arc where the lens boundary appears.

SE Channel Recalibration. After spatial gating, a Squeeze-and-Excitation (SE) block [20] recalibrates $\boldsymbol{F}^{*}$ channel-wise with reduction ratio r=16:

$$\boldsymbol{F}^{**} = \boldsymbol{F}^{*} \ \odot\ \sigma\Big(\mathbf{W}_2\,\mathrm{ReLU}\big(\mathbf{W}_1\,\mathrm{GAP}(\boldsymbol{F}^{*})\big)\Big), \quad (7)$$

where GAP denotes global average pooling. MGSA gates the spatial (H×W) dimensions of the C-channel feature map, zeroing activations outside the iris annulus uniformly across all channels; SE then reweights those same C channels globally, selecting which feature types encode the lens signal most strongly. The two are complementary: MGSA determines where to look; SE determines what to look for within that region.

3.5. Training and Testing

Subject-stratified cross-validation. Biometric datasets exhibit strong intra-subject correlation: all images of the same identity share iris texture, colour, and periocular appearance independent of lens type. Naive random splitting allows the same subject to appear on both sides of a fold boundary, enabling the model to exploit identity-specific cues rather than lens appearance—a form of *identity leakage* that inflates validation accuracy and masks poor generalization. We eliminate this with subject-stratified 5-fold cross-validation. (i) *identity disjointness*—all images of the same subject are confined to a single fold; and (ii) *class-balance preservation*—the three-class distribution is maintained across folds. The held-out test split (20% of subjects, never seen during training or model selection) is fixed at the dataset level. The best model checkpoint for each fold is selected by balanced accuracy on that fold's held-out training subjects. Here "checkpoint" denotes the saved model weights from the best-performing epoch during training. "Held-out training subjects" are the subjects kept aside from the training fold as the validation data to decide which epoch is best — this is different from the final test set. "Balanced accuracy" is used instead of just "accuracy" because the two classes (clear vs. normal) may not have equal numbers of images. Balanced accuracy averages the per-class recall, so a model that ignores the minority class cannot score well.

Ensemble inference. At test time, the five independently trained Stage-2 models $\{f_k\}_{k=1}^{5}$ are combined by averaging their softmax posterior distributions:

$$\hat{p}(c \mid I, M) = \frac{1}{5}\sum_{k=1}^{5} p_k(c \mid I, M),\ \ \hat{y} = \arg\max_{c} \hat{p}(c \mid I, M). \quad (8)$$

Since each fold is trained on a disjoint subject population, the five models are complementary in the identities they have observed; averaging their posteriors reduces decision-boundary variance without any additional computational cost at training time. For the optimization, stage 2 is trained with AdamW [41] (λ=10^{-4}), cosine annealing (η_0=2×10^{-4}, $\eta_{\min}$=10^{-6}, T=80 epochs), cross-entropy loss with label smoothing (ε=0.05), and dropout (p=0.4) before the classification head. Stage 1 is fine-tuned with SGD (η_0=5×10^{-4}, step decay $\times0.1$ every 10 epochs, 50 epochs total, weight decay 10^{-4}).

3.6. Lens-informed Score Calibration

Once the two-stage classifier assigns a label to every enrolled and probe image, we apply a per-condition z-score calibration to the raw VeriEye verification scores. For each pair type $t \in \{\texttt{NN}, \texttt{CC}, \texttt{CN}\}$, we estimate the mean μ_t and standard deviation σ_t from the *genuine* pairs of that type in the training set. Every score is then rescaled onto the Normal-Normal (NN) reference distribution:

$$s_{\text{cal}} = \frac{s - \mu_t}{\sigma_t}\,\sigma_{\text{NN}} + \mu_{\text{NN}}, \tag{9}$$

where, $(\mu_{\text{NN}}, \sigma_{\text{NN}})$ are the parameters of the NN reference distribution of the training set. This mapping is a structure-preserving affine transform: it shifts the mean of each non-NN distribution onto the NN mean and matches its spread, closing the score gap introduced by the lens without altering the rank ordering within any single pair-type group. Crucially, the pair type for each verification pair is determined entirely by the classifier's *predicted* labels—no ground-truth lens information is used at inference time.[3]

4. Experiments and Results

4.1. Contact Lens Classification

Table 1 reports Stage-1, Stage-2, and full pipeline accuracy together with per-class scores across all four datasets. Table 2 provides the per-class correct-prediction breakdown. Stage-1 achieves perfect patterned detection. Across all four datasets, Stage-1 (D-NetPAD) correctly classifies every patterned iris in the test set — 100% accuracy with zero bootstrap SD (Table 2). Patterned lenses introduce salient colour and periodic texture artefacts that are easily separable from natural iris textures, confirming that binary patterned/non-patterned classification is a well-conditioned problem for a standard convolutional network fine-tuned from ImageNet weights. Clear lens detection is the harder sub-problem. The main classification challenge lies in Stage-2. Across all datasets, the Clear lens accuracy is the lowest of the three classes, confirming that clear lens detection is harder than patterned detection and also harder than normal iris detection. As shown in Table 1, the gap is most pronounced on AD100 ($\text{S2}_{\text{Acc.}}$ = 87.50%) and LG4000 ($\text{S2}_{\text{Acc.}}$ = 92.00), which are the smallest datasets: with fewer training subjects, the model has less opportunity to learn the subtle reflectance deviation at the limbal boundary that distinguishes a clear lens from a bare iris. MGSA drives good Stage-2 accuracy on large datasets. On the two large IITD datasets, the MGSA-equipped ConvNeXt-Base achieves 96.15% (Cogent) and 98.18% (Vista) Stage-2 accuracy. By hard-zeroing all activations outside the iris and periocular arc via the anatomical mask $\tilde{M}$, the network cannot exploit background structures; the soft learned gate G then refines attention within that constrained region to the specific annular sub-areas where lens evidence physically appears. For cross-sensor analysis, despite substantial differences in sensor type, image resolution, and subject demographics across the four datasets, the pipeline achieves 90.0%–98.8% accuracy with a bootstrap SD between 0.45 and 4.23. This cross-sensor consistency supports the generalization of the proposed two-stage method.

4.2. Score and Calibration

Clear lenses cause systematic score degradation. Fig. 3 visualizes the VeriEye genuine-pair score distributions for each pair type before calibration. In every dataset, Clear-Clear (CC) genuine pairs score measurably lower than Normal-Normal (NN) pairs on the same subjects. The score gap ranges from 36 points on AD100 to 39 points on IITD-Vista, consistent with the systematic bias introduced by the refractive index mismatch at the lens boundary. Clear–Normal (CN) cross-condition pairs exhibit an even larger displacement on IITD-Vista (μ_{CN} = 650 vs. μ_{NN} = 830), pushing genuine match scores closer to the impostor distribution (grey curve, concentrated near zero). This confirms that clear lenses introduce a *predictable, condition-dependent* bias in iris matching–not random noise–and that the bias is present across all four sensors and cohorts tested. Calibration closes the gap. After applying the z-score calibration of Eq. (9) with pair types predicted by the two-stage classifier, the CC and CN distributions shift to align with the NN reference (Fig. 3, solid curves). On IITD-Vista, the CN mean rises from 650 to 830, completely closing the 180-point gap. On AD100, the CC distribution (amber solid) overlaps almost perfectly with the NN reference (blue), confirming that the calibration parameters estimated from genuine pairs generalise correctly to all pairs in the verification database. Crucially, this improvement is achieved using only the *predicted* lens types from the two-stage classifier–no ground-truth label information is available at inference time. Verification EER improves across all datasets. Table 3 and Fig. 4 report EER before and after calibration. The improvement is consistent across all four datasets, ranging from a modest 4.1% relative reduction on

[3] We also performed calibration using ground-truth lens labels and obtained the same accuracy, confirming that the predicted labels were sufficient for this evaluation.

Table 1. Two-stage contact-lens detection results across all four datasets. Stage 1 (D-NetPAD) classifies patterned vs. non-patterned irises. Stage 2 (ConvNeXt-Base + MGSA) classifies normal vs. clear irises on the non-patterned subset; therefore, Stage 2 accuracy is computed only on non-patterned test images. Accuracy ± values denote bootstrap standard errors from 10,000 resamples. Train/test splits are subject-stratified with no identity overlap. *Approximately 50 images with apparent labeling inconsistencies were excluded from the UND and IITD datasets during preprocessing; however, these inconsistencies were not exhaustively verified across the full datasets. [b]Baseline CCR (%) reproduced from ContlensNet [17] using intra-sensor validation from Table 1. ContlensNet was trained and tested using its own IIITD/ND protocol and a different number of images; therefore, its results are not directly comparable to the subject-stratified splits used here.

Dataset	**Dataset Statistics***					**Accuracy (%)**				**Subject-Stratified Split**			
	Imgs	**IDs**	**Pat.**	**Nor.**	**Clr.**	**S1**	**S2**	**Pipe.**	**ContlensNet[b]**	**Tr. IDs**	**Tr. N**	**Te. IDs**	**Te. N**
UND-AD100 [42]	300	41	100	100	100	100.00	87.50 ± 5.28	$\mathbf{90.00 \pm 4.23}$	95.00	32	250	9	50
UND-LG4000 [42]	1200	76	400	400	400	100.00	92.00 ± 2.22	$\mathbf{95.20 \pm 1.35}$	96.91	61	950	15	250
IITD-Cogent [43, 13]	3508	101	1202	1163	1143	100.00	96.15 ± 0.87	$\mathbf{97.50 \pm 0.56}$	86.73	79	2747	22	761
IITD-Vista [43, 13]	2906	101	1005	940	961	100.00	98.18 ± 0.69	$\mathbf{98.80 \pm 0.45}$	87.33	81	2322	20	584
Average	–	–	–	–	–	**100.00**	**93.46**	**95.38**	91.49	–	–	–	–

S1: Stage 1; *S2*: Stage 2 evaluated on the non-patterned subset only; *Pat.*: patterned; *Nor.*: normal; *Clr.*: clear; *Pipe.*: complete pipeline; *Tr.*: training split; *Te.*: test split.

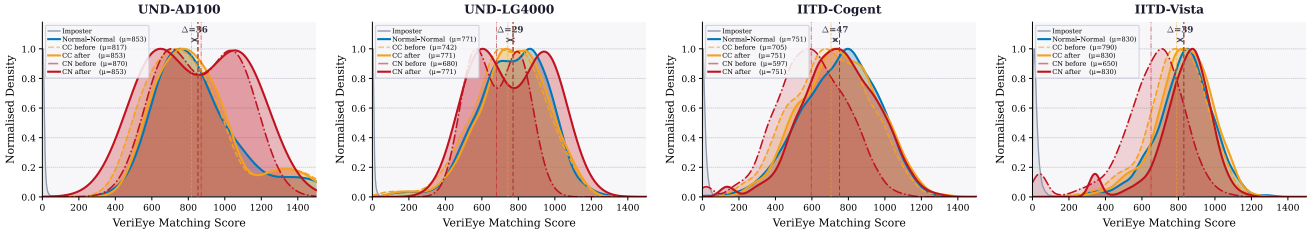

Figure 3. VeriEye genuine-pair score distributions before (dashed/faded) and after (solid) lens-informed score calibration, for all four datasets. **Blue**: Normal–Normal (reference, unchanged). **Amber**: Clear–Clear before → after. **Red**: Clear–Normal before → after. The Δ annotation marks the mean gap between Normal–Normal and Clear–Clear before calibration. Pair types are determined by the two-stage classifier predictions (no ground-truth labels used).

Table 2. Per-class correct predictions on a test split (correct / total). "Total" is the number of test images, and "correct" is the correctly classified test images.

Dataset	**Patterned**	**Normal**	**Clear**	**Overall**
UND-AD100	10/10 (100.0%)	16/20 (80.0%)	19/20 (95.0%)	**45/50 (90.00%)**
UND-LG4000	100/100 (100.0%)	80/80 (100.0%)	58/70 (82.9%)	**238/250 (95.20%)**
IITD-Cogent	268/268 (100.0%)	240/243 (98.8%)	234/250 (93.6%)	**742/761 (97.50%)**
IITD-Vista	200/200 (100.0%)	192/192 (100.0%)	185/192 (96.4%)	**577/584 (98.80%)**

AD100 (0.49% → 0.47%) to a substantial 28.30% on IITD-Vista (2.26% → 1.62%). The magnitude of the EER improvement correlates with the severity of the initial score degradation: IITD-Vista has the largest gap (39 points) and the largest CN displacement, and accordingly shows the greatest benefit from calibration. AD100, with a smaller Δ (36 points) and a small number of non-patterned pairs, shows the most modest improvement. The ROC curves in Fig. 4 confirm that the improvement is not confined to the

Table 3. VeriEye EER (%) before and after score calibration. Calibration parameters are estimated from genuine and impostor pairs; the presence or absence of a lens is determined by the proposed two-stage classifier (real-world protocol, no oracle labels used).

Dataset	**Before (%)**	**After (%)**	**Rel. Red.**
UND-AD100	0.49	0.47	−4.1%
UND-LG4000	0.75	0.64	−14.7%
IITD-Cogent	0.76	0.70	−7.9%
IITD-Vista	2.26	1.62	−28.3%

EER operating point: the calibrated curve (solid) lies above the raw curve (dashed) across the full range of False Match Rates, from 10^{-4} to 10^{0}. This means that regardless of the security threshold chosen by the operator, calibrated scores yield a higher TMR than raw scores.

4.3. Backbone Comparison

Table 4 compares three Stage-2 backbone architectures. ResNet-101, EfficientNet-V2-S, and ConvNeXt-Base all incorporate MGSA spatial attention. MGSA substantially outperforms the mask-free baseline. We conducted experiments on EfficientNet-V2-S without mask guidance, which degraded by up to 30% on AD100 (60.00% vs.

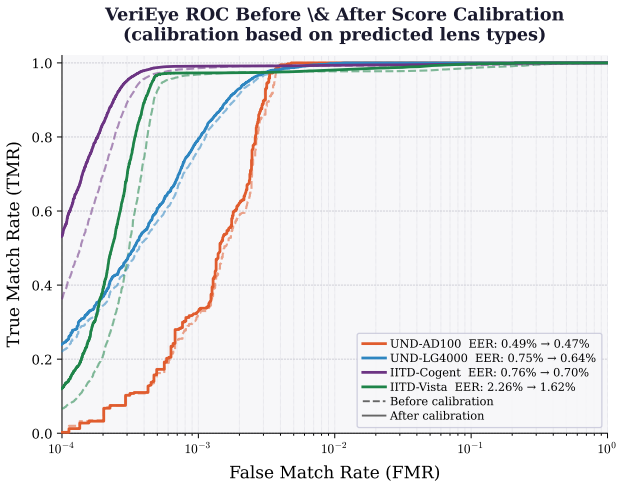

Figure 4. VeriEye ROC curves before (dashed) and after (solid) score calibration for all four datasets. EER values are annotated in the legend. Calibration uses pair types predicted by the proposed two-stage classifier under a real-world protocol (no oracle labels, 100% pair coverage).

Table 4. Stage-2 backbone comparison: full pipeline accuracy (%) with standard deviations from 10,000 resamples. Stage 1 is fixed.

Dataset	EfficientNet-V2 + MGSA	ResNet-101 + MGSA	ConvNeXt-B + MGSA
UND-AD100	58.00 ± 6.99	90.00 ± 4.26	**90.00 ± 4.26**
UND-LG4000	72.00 ± 2.83	**95.60 ± 1.29**	95.20 ± 1.38
IITD-Cogent	68.20 ± 1.67	97.24 ± 0.60	**97.50 ± 0.56**
IITD-Vista	65.41 ± 1.96	98.29 ± 0.54	**98.80 ± 0.45**

Table 5. Cross-dataset pipeline accuracy (%). Rows = train dataset; columns = test dataset. †Cross-sensor only (shared subjects).

Train ↓ Test →	UND-AD100	UND-LG4000	IITD-Cogent	IITD-Vista
UND-AD100	—	87.60	69.12	36.99
UND-LG4000	68.00	—	61.63	43.66
IITD-Cogent	78.00	**97.20**	—	†**97.26**
IITD-Vista	84.00	73.20	†85.41	—

90.00%) and by 10 points on IITD-Cogent (87.25% vs. 97.50%), confirming that the anatomical mask is the primary driver of Stage-2 performance, not backbone capacity alone. ConvNeXt-Base is the best MGSA backbone. Under identical mask guidance, ConvNeXt-Base matches or exceeds ResNet-101 on three of four datasets; ResNet-101 holds a marginal edge on LG4000 (+0.40 percentage points (pp)). On the larger IITD datasets, ConvNeXt-Base leads by +0.26 pp on Cogent and +0.51 pp on Vista, indicating that its 7×7 depthwise kernels better capture the subtle limbal texture evidence of clear lens wear. The 7×7 depthwise kernels of ConvNeXt capture the spatially extended limbal arc more effectively than ResNet's 3×3 convolutions, and this advantage is most pronounced where there is sufficient data to exploit the larger receptive field.

4.4. Cross-Dataset Generalization

Table 5 reports pipeline accuracy for all 12 source→target combinations using the same checkpoints as Table 1, with no fine-tuning on the target dataset. The results show that the proposed method generalizes well within the same sensor family and moderately across related sensors, with the IITD-Cogent model achieving 97.20% on UND-LG4000 and 97.26% on IITD-Vista, suggesting that the anatomical mask provides a robust, sensor-invariant spatial prior. Even under cross-family transfer, IITD-trained models remain well above chance, indicating that MGSA captures meaningful limbal cues rather than purely sensor-specific patterns. In contrast, models trained on the smaller UND datasets generalize poorly to IITD sensors, highlighting an asymmetric transfer gap likely caused by limited training diversity and substantial differences in image resolution and depth of field, and motivating future work on domain adaptation and stronger data augmentation.

5. Conclusion

We presented a two-stage method for iris contact lens detection that exploits the visual asymmetry between patterned and clear lenses. Stage 1 addresses the easier problem of patterned lens detection from cropped iris images with high accuracy across all four datasets, while Stage 2 addresses the harder clear vs. normal classification problem using a ConvNeXt-Base model with the proposed MGSA, which combines an anatomical ROI mask, spatial gating, and channel recalibration to focus on the most informative iris and periocular regions. Across UND-AD100, UND-LG4000, IITD-Cogent, and IITD-Vista, the full pipeline achieves 90.0%–98.8% accuracy and bootstrap SD scores, with ConvNeXt-Base consistently outperforming EfficientNet and ResNet baselines. We further showed that clear lenses systematically degrade commercial iris matcher scores, pushing genuine scores closer to the impostor distribution, and that condition-aware z-score calibration based on the predicted lens labels can recover much of this loss, reducing EER by 4.1%–28.3% across all datasets. Although the current system relies on Hough-based anatomical masks and has been evaluated only on near-infrared data, the results demonstrate that lens-aware detection and calibration provide a practical path toward more robust real-world iris recognition.

Acknowledgments: This work was supported by NSF CITeR under Awards #1841517 and #2413309. The code supporting the findings of this study is publicly available on GitHub [https://github.com/iPRoBe-lab/MGSA_ContactLens_Detection].

References

[1] J. Daugman and C. Downing, "Epigenetic randomness, complexity and singularity of human iris patterns," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1477, pp. 1737–1740, 2001. 1

[2] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004. 1, 2

[3] G. Erdogan and A. Ross, "Automatic detection of non-cosmetic soft contact lenses in ocular images," in *Biometric and Surveillance Technology for Human and Activity Identification X*, vol. 8712. SPIE, 2013, pp. 62–76. 1

[4] S&S Insider, "Contact Lens Market Size, Share & Growth Report 2035," https://www.snsinsider.com/reports/contact-lens-market-8566, 2026, last updated: March 2, 2026. Accessed: April 30, 2026. 1

[5] A. Boyd, J. Speth, L. Parzianello, K. W. Bowyer, and A. Czajka, "Comprehensive study in open-set iris presentation attack detection," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 3238–3250, 2023. 2

[6] S. Hoffman, R. Sharma, and A. Ross, "Convolutional neural networks for iris presentation attack detection: Toward cross-dataset and cross-sensor generalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018. 2

[7] D. Yambay, P. Das, A. Boyd, J. McGrath, Z. Fang, A. Czajka, S. Schuckers, K. Bowyer, M. Vatsa, R. Singh *et al.*, "Review of iris presentation attack detection competitions," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*. Springer, 2023, pp. 149–169. 2

[8] A. Boyd, Z. Fang, A. Czajka, and K. W. Bowyer, "Iris presentation attack detection: Where are we now?" *Pattern Recognition Letters*, vol. 138, pp. 483–489, 2020. 2

[9] S. Yadav and A. Ross, "CIT-GAN: Cyclic image translation generative adversarial network with application in iris presentation attack detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2412–2421. 2

[10] ——, "A multi-domain image translative diffusion stylegan for iris presentation attack detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop*, 2025, pp. 3688–3697. 2

[11] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcão, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, 2015. 2, 3

[12] M. Gupta, V. Singh, A. Agarwal, M. Vatsa, and R. Singh, "Generalized iris presentation attack detection algorithm under cross-database settings," in *25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 5318–5325. 2

[13] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer, "Unraveling the effect of textured contact lenses on iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 851–862, 2014. 2, 7

[14] S. Baker, K. W. Bowyer, and P. J. Flynn, "Degradation of iris recognition performance due to non-cosmetic prescription contact lenses," *Computer Vision and Image Understanding*, vol. 114, no. 9, pp. 1030–1044, 2010. 2, 3

[15] J. S. Doyle, K. W. Bowyer, and P. J. Flynn, "Variation in accuracy of textured contact lens detection based on sensor and lens pattern," in *IEEE Sixth International Conference on Biometrics: Theory, applications and Systems (BTAS)*, 2013, pp. 1–7. 2, 3

[16] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer, "Unraveling the effect of textured contact lenses on iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 851–862, 2014. 2, 3

[17] R. Raghavendra, K. B. Raja, and C. Busch, "Contlensnet: Robust iris contact lens detection using deep convolutional neural networks," in *2017 IEEE winter conference on applications of computer vision (WACV)*, pp. 1160–1167. 2, 3, 7

[18] K. Nguyen, H. Proença, and F. Alonso-Fernandez, "Deep learning for iris recognition: A survey," *ACM Computing Surveys*, vol. 56, no. 9, pp. 1–35, 2024. 2

[19] R. Sharma and A. Ross, "D-NetPAD: An explainable and interpretable iris presentation attack detector," in *IEEE International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–10. 2, 3, 4

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141. 2, 3, 5

[21] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114. 2, 3, 5

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. 2, 3, 5

[23] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 976–11 986. 2, 3, 5

[24] Neurotechnology, "VeriEye SDK," https://www.neurotechnology.com/verieye.html, 2026, iris identification technology and software development kit. 2

[25] A. Czajka, Z. Fang, and K. Bowyer, "Iris presentation attack detection based on photometric stereo features," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 877–885. 2

[26] A. Ross, S. Banerjee, C. Chen, A. Chowdhury, V. Mirjalili, R. Sharma, T. Swearingen, and S. Yadav, "Some research problems in biometrics: The future beckons," in *International Conference on Biometrics (ICB)*, 2019, pp. 1–8. 2

[27] A. Agarwal, N. Ratha, A. Noore, R. Singh, and M. Vatsa, "Misclassifications of contact lens iris PAD algorithms: Is it gender bias or environmental conditions?" in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 961–970. 3

[28] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708. 3, 4

[29] P. Das, J. McGrath, Z. Fang, L. Parzianello, and A. Czajka, "Iris liveness detection competition (LivDet-Iris) — the 2020 edition," in *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–10. 3

[30] C. Chen and A. Ross, "An explainable attention-guided iris presentation attack detector," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 2021, pp. 97–106. 3

[31] D. Yambay, B. Becker, N. Kohli, D. Yadav, A. Czajka, K. W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore, D. Gragnaniello, C. Sansone, and L. Verdoliva, "LivDet iris 2017 — iris liveness detection competition 2017," in *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 733–741. 3

[32] P. Tinsley, S. Purnapatra, M. Mitcheff, A. Boyd, C. Crum, K. Bowyer, P. Flynn, S. Schuckers, A. Czajka, M. Fang *et al.*, "Iris liveness detection competition (livdet-iris)–the 2023 edition," in *IEEE International Joint Conference on Biometrics (IJCB)*, 2023, pp. 1–10. 3

[33] M. Mitcheff, A. Hossain, S. Webster, S. Khan, K. Roszczewska, J. E. Tapia, F. Stockhardt, L. J. González-Soler, J.-Y. Lim, M. Pollok *et al.*, "Iris liveness detection competition (livdet-iris)–the 2025 edition," in *IEEE International Joint Conference on Biometrics (IJCB)*, 2025, pp. 1–10. 3

[34] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19. 3

[35] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803. 3

[36] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations (ICLR)*, 2021. 3

[37] B. Jafrasteh, S. P. Lubián-López, E. Trimarco, M. R. Ruiz, C. R. Barrios, Y. M. Almagro, and I. Benavente-Fernández, "Mga-net: A novel mask-guided attention neural network for precision neonatal brain imaging," *NeuroImage*, vol. 300, p. 120872, 2024. 3

[38] S. Shah and A. Ross, "Iris segmentation using geodesic active contours," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 824–836, 2009. 4

[39] G. W. Quinn, "An open source iris segmentation algorithm for non-ideal images," *NIST*, 2024. 4

[40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. 5

[41] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019. 6

[42] J. Doyle and K. W. Bowyer, "Notre dame image database for contact lens detection in iris recognition–2013," University of Notre Dame, Computer Vision Research Laboratory, Tech. Rep., 2014, dataset documentation/README. 7

[43] N. Kohli, D. Yadav, M. Vatsa, and R. Singh, "Revisiting iris recognition with color cosmetic contact lenses," in *Proceedings of the International Conference on Biometrics*, 2013, pp. 1–7. 7