Single-Image Pupil Localization via Implicit 3D Eye Reconstruction

Taejun Roh*, Yejin Cho†, Duong Hai Nguyen‡, and Chul Lee†‡
* Division of Electronics and Electrical Engineering, Dongguk University, Seoul, Korea
† Department of Computer Science and Artificial Intelligence, Dongguk University, Seoul, Korea
‡ Department of Multimedia Engineering, Dongguk University, Seoul, Korea
E-mails: {rohtj_312, yjcho2912}@dgu.ac.kr, duongnguyen@mme.dongguk.edu, and chullee@dongguk.edu

***Abstract*—We propose a pupil localization algorithm that directly estimates the pupil region in a single image based on geometric 3D priors and implicit 3D eye reconstruction using self-supervised learning. First, we develop a 3D eye reconstruction network that implicitly constructs a biologically inspired eye model from a single image by estimating geometric eye priors. Then, we project the reconstructed 3D eye model back onto the original image plane by developing an inverse ray tracing technique to localize the 2D pupil region. Since this projection is non-differentiable and 3D annotations are unavailable, we develop a self-supervised learning strategy that generates 3D pseudo-annotations from 2D pupil ground truths to train the reconstruction network. Experimental results demonstrate that the proposed algorithm achieves better performance than state-of-the-art algorithms in both quantitative and qualitative evaluations.**

I. Introduction

Pupil localization, a fundamental task that aims to accurately identify the pupil region in eye images, is used in a wide range of applications such as gaze estimation [1], human-computer interaction [2], psychological computing [3], and online education [4]. In addition, variations in pupil location and size serve as nonverbal indicators of cognitive and physiological states, providing a reliable basis for assessment in human-centered applications [5], [6]. Due to its importance in practical applications, pupil localization has recently received considerable attention.

Conventional pupil localization algorithms can be broadly categorized into image appearance-based and eye geometry-based approaches. Image appearance-based approaches estimate the pupil location directly from visual cues, as illustrated in Fig. 1(a). In particular, traditional model-based algorithms formulate pupil localization as an edge extraction followed by an ellipse fitting problem [7], [8]. While they are effective when the pupil boundary is clearly distinguishable, their performance may deteriorate under low-contrast conditions or in the presence of occlusions such as specular reflections.

This work was supported in part by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5C3A0104366711) and in part by the Ministry of Science and ICT, Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) in 2025 (2023-0-00049).

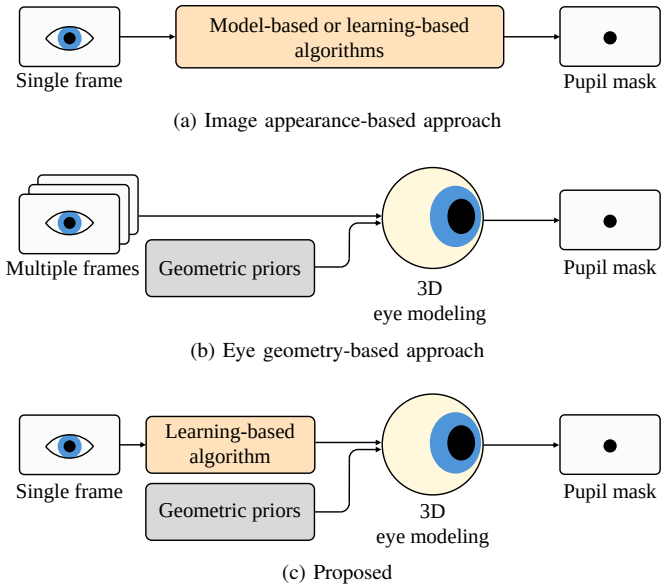


Fig. 1: Comparison of pupil localization approaches. (a) Image appearance-based approaches directly estimate the pupil region using 2D image cues. (b) Eye geometry-based approaches adopt 3D eye models. (c) The proposed algorithm reconstructs a 3D eye model from a single image using geometric priors.

In contrast, learning-based approaches [9]–[11] typically perform pixel-wise segmentation followed by ellipse fitting to determine the pupil region. Although these algorithms benefit from data-driven feature learning, they often struggle when the pupil is partially occluded by the eyelid or eyelashes, since such occlusions cause ambiguity in both the segmentation and contour fitting stages.

Eye geometry-based approaches in Fig. 1(b) adopt 3D eye models to overcome the limitations of image appearance-based approaches and improve robustness to occlusions and distortions. By incorporating geometric priors following 2D segmentation, they explicitly model the 3D structure of the eye and project it back onto the image plane to localize the pupil [12], [13]. However, such algorithms typically require temporal information across multiple frames or custom head-mounted camera systems for accurate 3D reconstruction, limiting their applicability in practical scenarios where only a

979-8-3315-7206-8/25/$31.00

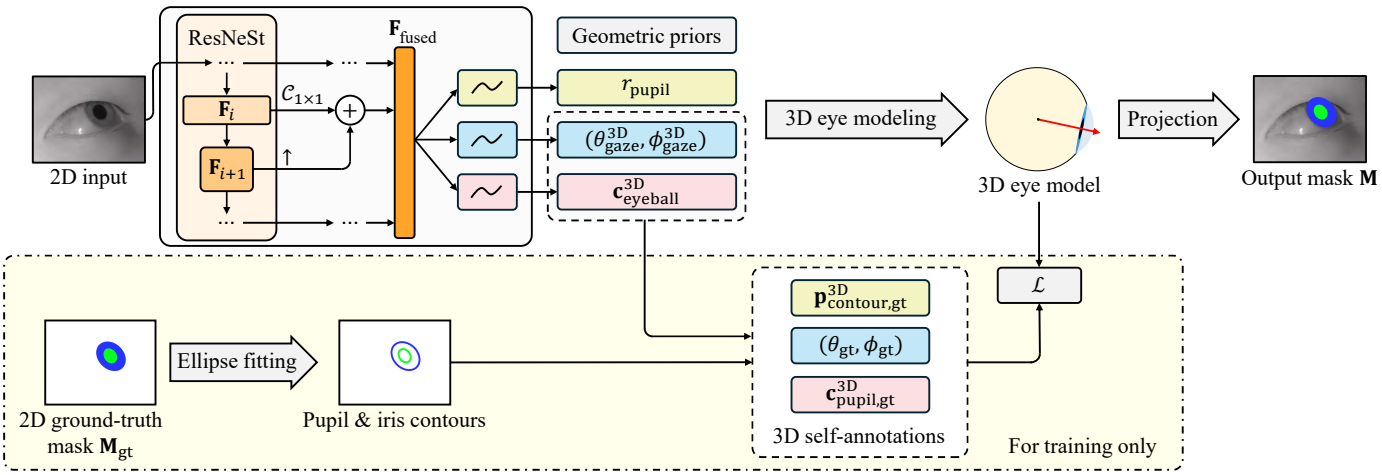


Fig. 2: Overview of the proposed algorithm. During inference, only geometric priors estimated from a single input eye image are required to reconstruct the 3D eye model and localize the pupil.

single frame is available. Furthermore, their dependence on 2D segmentation makes them susceptible to errors caused by occlusions and low-contrast regions, which significantly reduces the reliability of subsequent 3D estimation.

To address the aforementioned limitations of conventional approaches, we propose a pupil localization algorithm based on implicit 3D eye reconstruction. First, a biologically inspired 3D eye model is reconstructed from a single input eye image by combining learned eye features and geometric 3D priors. The reconstructed 3D eye model is then projected back onto the image plane using inverse ray tracing to localize the pupil region. To this end, we develop a self-supervised learning strategy that synthesizes 3D pseudo-annotations. Experimental results demonstrate that the proposed algorithm achieves better performance than state-of-the-art algorithms in both quantitative and qualitative comparisons.

II. Proposed Algorithm

Fig. 2 illustrates an overview of the proposed pupil localization algorithm. Given a single input eye image, a biologically inspired 3D eye model is constructed by estimating eye parameters and geometric eye priors. The reconstructed 3D model is then projected onto the original image plane via inverse ray tracing to accurately generate the pupil mask. During training, since ground-truth 3D annotations are unavailable and the projection function is non-differentiable, we develop a self-annotation algorithm that enables self-supervision in the 3D space. During inference, the proposed algorithm reconstructs the 3D eye model from a single input image and localizes the pupil region without requiring any annotations. In the following subsections, we describe each component of the proposed algorithm in detail.

A. Implicit 3D Eye Modeling

Conventional image appearance-based algorithms often struggle with occlusions caused by eyelids and eyelashes, as they fail to consider the inherent 3D structure of the eye. To address this limitation, we reconstruct an implicit 3D eye model from a single input image and project it back onto the image plane for accurate pupil localization. Specifically, we estimate the eyeball center $\mathbf{c}^{\text{3D}}_{\text{eyeball}}$, gaze direction $\mathbf{g}_{\text{angular}}$, and pupil radius r_{pupil}, which is a minimal yet sufficient set of parameters for reconstructing the eye geometry based on a biologically inspired model [12], [14], as illustrated in Fig. 3. In this model, the radii of the eyeball r_{eyeball}, corneal sphere r_{cornea}, and 3D iris contour r_{iris} are fixed and serve as geometric priors [15]–[17].

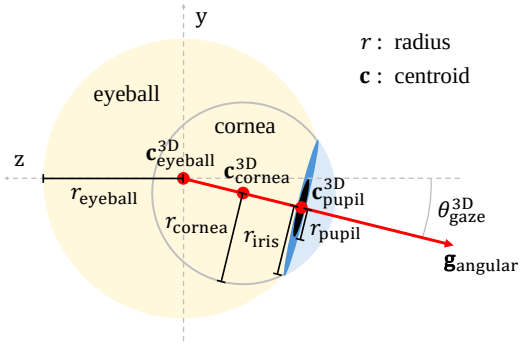


Fig. 3: Illustration of the biologically inspired 3D eye model [14].

We adopt ResNeSt [18] as the backbone of the 3D eye reconstruction network, as shown in Fig. 2. We leverage its split-attention blocks to capture both global context and fine-grained details to facilitate effective feature extraction from eye regions that exhibit significant variation in shape and size under uncontrolled imaging conditions. Specifically, we construct an L-level feature pyramid $\{\mathbf{F}_i\}_{i=1}^{L}$ from the input image and then apply a feature modulation mechanism [19] to retain both spatial details and high-level semantic information. More specifically, each feature map $\mathbf{F}_i$ is modulated by adding the

upsampled feature map from the previous level as

$$\mathbf{F}'_i = \mathcal{C}_{1\times1}(\mathbf{F}_i) + \uparrow \mathcal{C}_{1\times1}(\mathbf{F}_{i+1}), \tag{1}$$

where $\mathcal{C}_{1\times1}$ denotes a 1×1 convolution layer applied along the channel dimension and $\uparrow$ represents nearest-neighbor upsampling along the spatial dimension. The resulting modulated feature maps are then concatenated along the channel dimension to form a multiscale feature representation shared across multiple tasks—eyeball center, gaze direction, and pupil radius estimation—for 3D eye modeling, *i.e.*,

$$\mathbf{F}_{\text{fused}} = \left[..., \text{GAP}(\mathbf{F}'_i), \text{GAP}(\mathbf{F}'_{i+1}), ...\right], \tag{2}$$

where $\text{GAP}(\cdot)$ denotes a global average pooling layer [20]. Then, we estimate the geometric priors, *i.e.*, the parameters of 3D eyeball, gaze vector, and 3D pupil, using three dedicated nonlinear functions. This separation of coordinate, directional, and scalar parameters reduces mutual interference and thus improves both training stability and prediction accuracy. In this work, each nonlinear function is implemented as a multilayer perceptron with a single hidden layer.

Next, we compute the centers of the corneal sphere $\mathbf{c}^{\text{3D}}_{\text{cornea}}$ and the pupil/iris circle $\mathbf{c}^{\text{3D}}_{\text{pupil}}$ for the eye model in Fig. 3 using the estimated geometric priors. The corneal center $\mathbf{c}^{\text{3D}}_{\text{cornea}}$ lies along the gaze direction at a distance d_{cornea} from the eyeball center $\mathbf{c}^{\text{3D}}_{\text{eyeball}}$, while the pupil/iris center $\mathbf{c}^{\text{3D}}_{\text{pupil}}$ is located at a distance d_{pupil} from $\mathbf{c}^{\text{3D}}_{\text{eyeball}}$. These distances are computed as

$$\begin{aligned} d_{\text{cornea}} &= \sqrt{r^2_{\text{eyeball}} - r^2_{\text{iris}}} - \sqrt{r^2_{\text{cornea}} - r^2_{\text{iris}}}, \\ d_{\text{pupil}} &= \sqrt{r^2_{\text{eyeball}} - r^2_{\text{iris}}}. \end{aligned} \tag{3}$$

Thus, the 3D positions $\mathbf{c}^{\text{3D}}_{\text{cornea}}$ and $\mathbf{c}^{\text{3D}}_{\text{pupil}}$ are obtained by

$$\begin{aligned} \mathbf{c}^{\text{3D}}_{\text{cornea}} &= \mathbf{c}^{\text{3D}}_{\text{eyeball}} + d_{\text{cornea}} \cdot \mathbf{g}_{\text{cart}}, \\ \mathbf{c}^{\text{3D}}_{\text{pupil}} &= \mathbf{c}^{\text{3D}}_{\text{eyeball}} + d_{\text{pupil}} \cdot \mathbf{g}_{\text{cart}}, \end{aligned} \tag{4}$$

where $\mathbf{g}_{\text{cart}}$ is the Cartesian form of the gaze direction $\mathbf{g}_{\text{angular}} = \left[\theta^{\text{3D}}_{\text{gaze}}, \phi^{\text{3D}}_{\text{gaze}}\right]^\top$. Specifically, $\mathbf{g}_{\text{cart}}$ is given by

$$\mathbf{g}_{\text{cart}} = \left[\sin\theta^{\text{3D}}_{\text{gaze}} \cos\phi^{\text{3D}}_{\text{gaze}}, \sin\theta^{\text{3D}}_{\text{gaze}} \sin\phi^{\text{3D}}_{\text{gaze}}, -\cos\theta^{\text{3D}}_{\text{gaze}}\right]^\top. \tag{5}$$

B. Pupil and Iris Localization via Inverse Ray Tracing

To obtain a 2D binary pupil mask that represents the pupil region, we project the estimated 3D eye model onto the image plane. However, direct projection of the 3D eye model requires iterative optimization due to the nonlinear refraction of light at the curved corneal surface, which is computationally expensive and prone to inaccuracies. To address this challenge, we develop an inverse ray tracing strategy, as illustrated in Fig. 4. Specifically, for each 2D pixel $\mathbf{p}^{\text{2D}}$ in the binary mask $\mathbf{M}$, a ray $\mathbf{r}^{\text{3D}}$ is traced from $\mathbf{p}^{\text{2D}}$ through the camera center $\mathcal{O}$, and the corresponding 3D point $\mathbf{p}^{\text{3D}}_{\text{pupil}}$ is determined as the intersection of this ray with the pupil-iris plane. The complete inverse ray tracing process for computing $\mathbf{p}^{\text{3D}}_{\text{pupil}}$ is summarized in Algorithm 1. Next, we refine the binary mask

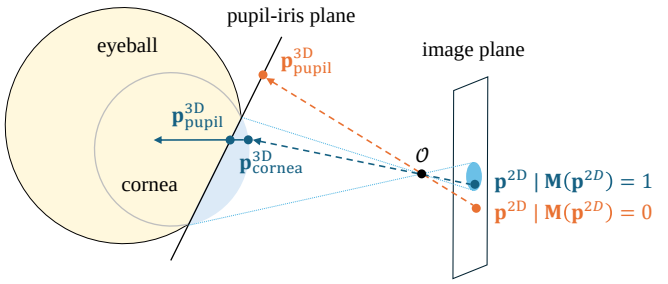


Fig. 4: Illustration of the inverse ray tracing. For a pixel $\mathbf{p}^{\text{2D}}$ in the image plane, the corresponding 3D point $\mathbf{p}^{\text{3D}}_{\text{pupil}}$ is determined as the intersection of the back-projected ray with the pupil-iris plane in the 3D camera coordinate system.

Algorithm 1 Inverse ray tracing $\text{Trace}_{\text{3D}}\left(\mathbf{p}^{\text{2D}}, \mathbf{c}^{\text{3D}}_{\text{pupil}}, \mathbf{g}_{\text{cart}}, f\right)$

Input: 2D pixel $\mathbf{p}^{\text{2D}} = \left[\text{p}^{\text{2D}}_x, \text{p}^{\text{2D}}_y\right]^\top$, pupil center $\mathbf{c}^{\text{3D}}_{\text{pupil}}$, gaze vector $\mathbf{g}_{\text{cart}}$, focal length f

Compute normalized ray direction $\mathbf{r}^{\text{3D}} \leftarrow \frac{1}{\sqrt{|\text{p}^{\text{2D}}_x|^2 + |\text{p}^{\text{2D}}_y|^2 + |f|^2}} \left[\text{p}^{\text{2D}}_x, \text{p}^{\text{2D}}_y, f\right]^\top$

if ray intersects corneal surface **then**

  $\mathbf{p}^{\text{3D}}_{\text{pupil}} \leftarrow$ Unprojection function [12]

else

  $\mathbf{p}^{\text{3D}}_{\text{pupil}} \leftarrow \frac{\left(\mathbf{c}^{\text{3D}}_{\text{pupil}}\right)^\top \mathbf{g}_{\text{cart}}}{\left(\mathbf{r}^{\text{3D}}\right)^\top \mathbf{g}_{\text{cart}}} \cdot \mathbf{r}^{\text{3D}}$

end if

Output: Unprojected 3D point $\mathbf{p}^{\text{3D}}_{\text{pupil}}$

$\mathbf{M}$ by evaluating the distance between $\mathbf{p}^{\text{3D}}_{\text{pupil}}$ and the 3D pupil center $\mathbf{c}^{\text{3D}}_{\text{pupil}}$ as

$$\mathbf{M}(\mathbf{p}^{\text{2D}}) = \begin{cases} 1, & \text{if } \left\|\mathbf{p}^{\text{3D}}_{\text{pupil}} - \mathbf{c}^{\text{3D}}_{\text{pupil}}\right\|_2 \le r_{\text{pupil}}, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The same procedure is applied to the iris by substituting the pupil radius r_{pupil} with the iris radius r_{iris}.

C. Self-Supervised Learning

We train the proposed algorithm using a loss function defined as the weighted sum of center loss $\mathcal{L}_{\text{center}}$, gaze loss $\mathcal{L}_{\text{gaze}}$, and contour loss $\mathcal{L}_{\text{contour}}$ as

$$\mathcal{L} = \lambda_{\text{center}}\mathcal{L}_{\text{center}} + \lambda_{\text{gaze}}\mathcal{L}_{\text{gaze}} + \lambda_{\text{contour}}\mathcal{L}_{\text{contour}}, \tag{7}$$

where λ_{center}, λ_{gaze}, and λ_{contour} are the weighting coefficients for each loss term. These losses are designed to supervise the 3D eye model reconstruction. However, since ground-truth 3D annotations are unavailable, we develop a self-labeling strategy that synthesizes 3D pseudo-annotations from the ground-truth 2D pupil and iris masks $\mathbf{M}_{\text{gt}}$. This enables self-supervised learning of the implicit 3D eye model parameters: the eyeball center $\mathbf{c}^{\text{3D}}_{\text{eyeball}}$, gaze vector $\left[\theta^{\text{3D}}_{\text{gaze}}, \phi^{\text{3D}}_{\text{gaze}}\right]^\top$, and pupil radius r_{pupil}. We subsequently detail each loss term and the corresponding self-annotation generation process.

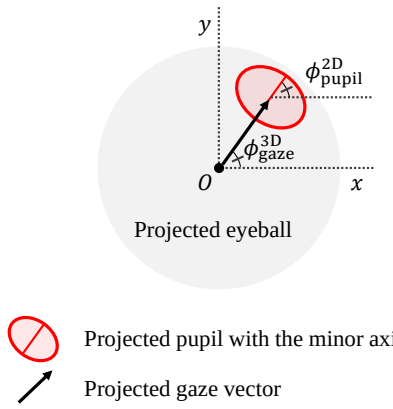


Fig. 5: Illustration of the gaze horizontal angle ϕ^{3D}_{gaze} and the in-plane rotation angle ϕ^{2D}_{pupil} of the projected iris ellipse.

1) Center Loss $\mathcal{L}_{center}$: We define the center loss as a logarithm function of the distance between the estimated 3D eyeball center $\mathbf{c}^{3D}_{eyeball}$ and the self-annotated 3D pupil center $\mathbf{c}^{3D}_{pupil,gt}$ as

$$\mathcal{L}_{center} = \log\left(1 + \left\|\mathbf{c}^{3D}_{pupil,gt} - \mathbf{c}^{3D}_{pupil}\right\|_2\right), \tag{8}$$

where $\mathbf{c}^{3D}_{pupil,gt}$ is inferred from the 2D pupil centroid $\mathbf{c}^{2D}_{pupil,gt}$ using inverse ray tracing in Algorithm 1 as

$$\mathbf{c}^{3D}_{pupil,gt} = \text{Trace}_{3D}\left(\mathbf{c}^{2D}_{pupil,gt}, \mathbf{c}^{3D}_{pupil}, \mathbf{g}_{cart}, f\right). \tag{9}$$

2) Gaze Loss $\mathcal{L}_{gaze}$: We define the gaze loss $\mathcal{L}_{gaze}$ as the angular error between the predicted gaze vector $\left(\theta^{3D}_{gaze}, \phi^{3D}_{gaze}\right)$ and the self-annotated ground-truth vector $\left(\theta_{gt}, \phi_{gt}\right)$. First, to obtain the horizontal gaze angle ϕ_{gt}, we fit an ellipse to the 2D iris contour in the ground-truth mask $\mathbf{M}_{gt}$ using ellipse fitting [21]. The resulting ellipse is represented by $(\mathbf{c}^{2D}_{iris}, a^{2D}_{iris}, b^{2D}_{iris}, \phi^{2D}_{iris})$, where $\mathbf{c}^{2D}_{iris}$ is the ellipse center, a^{2D}_{iris} and b^{2D}_{iris} are the major and minor axes, respectively, and ϕ^{2D}_{iris} is the in-plane rotation angle between the minor axis and the horizontal image axis, as illustrated in Fig. 5. Based on the observation that $\phi^{3D}_{gaze} \approx \phi^{2D}_{iris}$, we set $\phi_{gt} = \phi^{2D}_{iris}$.

Second, to generate the vertical gaze angle θ_{gt}, we first sample a candidate angle $\tilde{\theta} \in \Theta$, where Θ is a discrete set of possible vertical angles. For each $\tilde{\theta}$, we reconstruct the 3D eye model as described in Section II-A and obtain a corresponding iris mask $\mathbf{M}_{\tilde{\theta}}$ using inverse ray tracing in Section II-B. Then, we select the optimal angle $\tilde{\theta}$ as the one that maximizes the Dice similarity [22] between the estimated and ground-truth masks, *i.e.*,

$$\theta_{gt} = \arg\max_{\tilde{\theta}\in\Theta} \text{Dice}\left(\mathbf{M}_{\tilde{\theta}}, \mathbf{M}_{gt}\right). \tag{10}$$

Next, we compute the angular error α between two gaze vectors using the spherical cosine law as

$$\cos\alpha = \sin\theta_{gt}\sin\theta^{3D}_{gaze}\cos\left(\phi_{gt} - \phi^{3D}_{gaze}\right) + \cos\theta_{gt}\cos\theta^{3D}_{gaze}, \tag{11}$$

The gaze loss is then defined as

$$\mathcal{L}_{gaze} = 1 - \cos\alpha. \tag{12}$$

3) Contour Loss $\mathcal{L}_{contour}$: Based on the eye model [14], which assumes that the pupil is a circle lying on the pupil-iris plane and that the iris has a constant radius, we define the contour loss to incorporate geometric constraints and eliminate scale ambiguity. Specifically, we construct a discrete set of ground-truth points $\left\{\mathbf{p}^{3D}_{pupil,gt}\right\}$ on the 3D pupil contour, where each point $\mathbf{p}^{3D}_{pupil,gt}$ is obtained via inverse ray tracing as

$$\mathbf{p}^{3D}_{pupil,gt} = \text{Trace}_{3D}\left(\mathbf{p}^{2D}_{pupil,gt}, \mathbf{c}^{3D}_{pupil}, \mathbf{g}_{cart}, f\right) \tag{13}$$

given the sampled 2D pupil contour point $\mathbf{p}^{2D}_{pupil,gt}$. Similarly, we construct the set $\left\{\mathbf{p}^{3D}_{iris,gt}\right\}$ for the iris contour. Then, we define the contour loss, which quantifies the fitting errors for both the pupil and iris, as

$$\begin{aligned}\mathcal{L}_{contour} = &\sum_{\mathbf{p}^{3D}_{pupil,gt}} \log\left(1 + \left|r_{pupil} - \left\|\mathbf{p}^{3D}_{pupil,gt} - \mathbf{c}^{3D}_{pupil}\right\|_2\right|\right) \\ &+ \sum_{\mathbf{p}^{3D}_{iris,gt}} \log\left(1 + \left|r_{iris} - \left\|\mathbf{p}^{3D}_{iris,gt} - \mathbf{c}^{3D}_{pupil}\right\|_2\right|\right).\end{aligned} \tag{14}$$

Note that r_{iris} is a geometric prior that provides a strong constraint to resolve scale ambiguity.

III. Experimental Results

A. Implementation

We train the proposed algorithm end-to-end for 400 epochs using the Adam optimizer [23] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and a batch size of 16. The learning rate is set to 10^{-8}, and it is linearly increased to 10^{-4} over the first 10 epochs as a warm-up stage. Thereafter, it is multiplied by a factor of 0.8 if the validation performance does not improve for 20 consecutive epochs. The loss weights λ_{center}, λ_{gaze}, and $\lambda_{contour}$ in (7) are set to 0.02, 1.0, and 0.01, respectively. After 300 epochs, λ_{center} is decreased to 0.01 to reduce its influence. We use fixed anatomical parameters from the Le Grand eye model [12], [14].

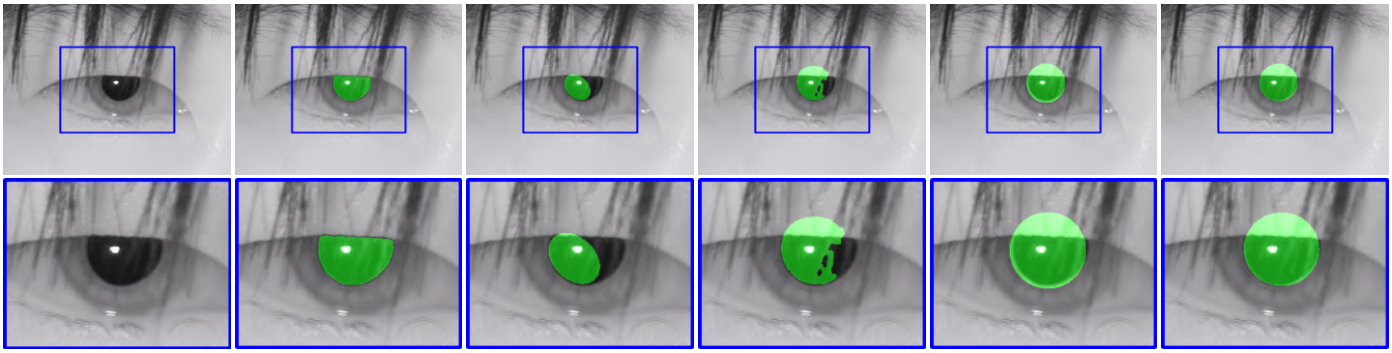
B. Quantitative and Qualitative Comparison

We evaluate the performance of the proposed algorithm against five state-of-the-art algorithms: Wahet [24], Iris-Seg [25], Iris-SAM [26], and the top two winners of the NIR-ISL Challenge [27]. For evaluation, we use five datasets from the NIR-ISL 2021 Challenge [27]—CASIA-Iris-Distance, CASIA-Iris-Occlusion, CASIA-Iris-Off-angle, CASIA-Iris-M1, and CASIA-Iris-Africa—all of which consist of single-channel near-infrared images. The Dice similarity coefficient [22] is used as the evaluation metric, where a higher score indicates better performance.

We present the results of the quantitative comparison of pupil localization performance in Table I. The proposed algorithm outperforms state-of-the-art algorithms in two out of five datasets—CASIA-Iris-Distance and CASIA-Iris-Occlusion—demonstrating its effectiveness in pupil localization, especially under occlusion due to the use of an implicit 3D eye model. In addition, it achieves the second-best performance on the

TABLE I: Quantitative comparison of pupil localization performance. The **boldfaced** and underlined numbers denote the best and the second-best results, respectively.

	CASIA-Distance	CASIA-Occlusion	CASIA-Off-angle	CASIA-M1	CASIA-Africa
Wahet [24]	0.8473	0.8731	0.9450	0.8424	0.5668
IrisSeg [25]	0.9255	0.8642	0.8949	0.7520	0.7934
Iris-SAM [26]	0.9186	0.9412	0.9470	0.8759	**0.9521**
NIR2021041402 [27]	0.9382	0.8401	0.9317	**0.9556**	0.9042
NIR2021030902 [27]	0.9345	0.9401	**0.9476**	0.9369	0.9214
Proposed	**0.9511**	**0.9533**	0.9255	0.9458	0.8717



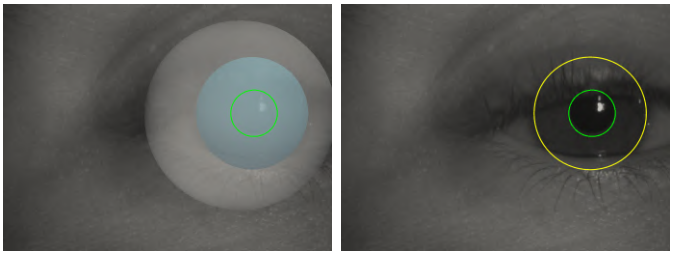
(a) Input (b) Iris-SAM (c) NIR2021041402 (d) NIR2021030902 (e) Proposed (f) Ground truth

Fig. 6: Qualitative comparison of the pupil localization results on the CASIA-Iris-Occlusion dataset. The bottom row shows the magnified views of the regions highlighted by blue rectangles in the top row.

CASIA-Iris-M1 dataset while maintaining comparable performance across the remaining datasets.

Fig. 6 shows the qualitative comparison of the pupil localization results of each algorithm on the CASIA-Iris-Occlusion dataset. Conventional algorithms, which rely on appearance-based features in the image domain, fail to capture the complete pupil structure under occlusion, *e.g.*, eyelids and eyelashes. For instance, Iris-SAM and NIR2021041402 in Fig. 6(b) and Fig. 6(c), respectively, fail to localize the pupil region occluded by the eyelid, resulting in noticeable shape distortions. Although NIR2021030902 in Fig. 6(d) successfully detects the pupil region beneath the eyelid, it fails to accurately identify the pupil in areas occluded by hair. In contrast, the proposed algorithm in Fig. 6(e) effectively captures the complete pupil region, even under severe occlusions, confirming the effectiveness of the implicit 3D eye model. These results indicate that the proposed algorithm is inherently more robust to occlusions due to the use of a higher-dimensional representation for modeling the eye.

Finally, Fig. 7(a) visualizes the implicit 3D eye model reconstructed by the proposed algorithm. The corresponding 2D iris and pupil contours in Fig. 7(b) are obtained by projecting the 3D model onto the image plane. The implicit 3D eye model enables the proposed algorithm to compute accurate pupil masks, which is not feasible for conventional algorithms that rely on appearance-based features, especially in the presence of occlusions.



(a) Predicted 3D eye model (b) Projected iris and pupil

Fig. 7: Visualization of the predicted 3D eye.

IV. Conclusions

We proposed a single-image pupil localization algorithm based on implicit 3D eye model reconstruction. Given a single eye image, we first constructed a biologically inspired 3D eye model by exploiting both implicit model parameters and geometric eye priors. We then projected the reconstructed model onto the original image plane to generate a complete pupil mask. Since the projection process is non-differentiable and 3D annotations are unavailable, we developed a self-supervised learning strategy to train the proposed algorithm. Specifically, 2D pupil ground truths are unprojected to generate 3D pseudo-annotations, which are then used to supervise the learning process. Experimental results demonstrated that the proposed algorithm achieves better performance than state-of-the-art algorithms.

References

[1] D. Hu and K. Huang, "Semi-supervised multitask learning using gaze focus for gaze estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, pp. 7935–7946, Sep. 2024.

[2] W. Zhang, M. L. Smith, L. N. Smith, and A. Farooq, "Eye center localization and gaze gesture recognition for human-computer interaction," *J. Opt. Soc. Am. A*, vol. 33, no. 3, pp. 314–325, Mar. 2016.

[3] G. Minadakis and K. Lohan, "Using pupil diameter to measure cognitive load," in *Proc. AAAI Fall Symp. AI-HRI*, Oct. 2018.

[4] H. Mitre-Hernandez, C. Carrillo Roberto, and L.-A. Carlos, "Pupillary responses for cognitive load measurement to classify difficulty levels in an educational video game: Empirical study," *JMIR Serious Games*, vol. 9, no. 1, Jan. 2021.

[5] Q. Zhuang, Z. Kehua, J. Wang, and Q. Chen, "Driver fatigue detection method based on eye states with pupil and iris segmentation," *IEEE Access*, vol. 8, pp. 173 440–173 449, Jan. 2020.

[6] T. Tabashum, A. Zaffer, R. Yousefzai, *et al.*, "Detection of Parkinson's disease through automated pupil tracking of the post-illumination pupillary response," *Front. Med.*, vol. 8, Mar. 2021.

[7] W. Fuhl, T. C. Santini, T. Kübler, and E. Kasneci, "ElSe: Ellipse selection for robust pupil detection in real-world environments," in *Proc. ACM Symp. Eye Track. Res. Appl.*, Mar. 2016, pp. 123–130.

[8] W. Fuhl, T. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci, "Excuse: Robust pupil detection in real-world scenarios," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, Jan. 2015, pp. 39–51.

[9] C. Akinlar, H. K. Kucukkartal, and C. Topal, "Accurate CNN-based pupil segmentation with an ellipse fit error regularization term," *Expert Syst. Appl.*, vol. 188, Feb. 2022.

[10] S. Y. Han, H. J. Kwon, Y. Kim, and N. I. Cho, "Noise-robust pupil center detection through CNN-based segmentation with shape-prior loss," *IEEE Access*, vol. 8, pp. 64 739–64 749, Apr. 2020.

[11] Y.-H. Yiu, M. Aboulatta, T. Raiser, *et al.*, "DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning," *J. Neurosci. Methods*, vol. 324, Jan. 2019.

[12] K. Dierkes, M. Kassner, and A. Bulling, "A novel approach to single camera, glint-free 3D eye model fitting including corneal refraction," in *Proc. ACM Symp. Eye Track. Res. Appl.*, Jun. 2018, pp. 1–9.

[13] K. Dierkes, M. Kassner, and A. Bulling, "A fast approach to refraction-aware eye-model fitting and gaze prediction," in *Proc. ACM Symp. Eye Track. Res. Appl.*, Jun. 2019, pp. 1–9.

[14] Y. L. Grand, *Light, Color and Vision*. New York, NY, USA: Wiley, 1957.

[15] I. Bekerman, P. Gottlieb, and M. Vaiman, "Variations in eyeball diameters of the healthy adults," *J. Ophthalmol.*, vol. 2014, no. 1, Nov. 2014.

[16] H. Gross, F. Blechinger, and B. Achtner, "Human eye," in *Handbook of Optical Systems*. Weinheim, Germany: John Wiley & Sons, 2008, ch. 36, pp. 1–87.

[17] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, Jun. 2006.

[18] H. Zhang, C. Wu, Z. Zhang, *et al.*, "ResNeSt: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jan. 2022, pp. 2736–2746.

[19] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2117–2125.

[20] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent.*, Apr. 2014.

[21] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, May 1999.

[22] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, May 2015.

[24] A. Uhl and P. Wild, "Weighted adaptive Hough and ellipsopolar transforms for real-time iris segmentation," in *Proc. IAPR Int. Conf. Biometrics*, Apr. 2012, pp. 283–290.

[25] A. Gangwar, A. Joshi, A. Singh, F. Alonso-Fernandez, and J. Bigun, "IrisSeg: A fast and robust iris segmentation framework for non-ideal iris images," in *Proc. Int. Conf. Biometrics*, Jun. 2016, pp. 1–8.

[26] P. Farmanifard and A. Ross, "Iris-SAM: Iris segmentation using a foundation model," in *Proc. Int. Conf. Pattern Recognit. Artif. Intell.*, Feb. 2024, pp. 394–409.

[27] C. Wang, Y. Wang, K. Zhang, *et al.*, "NIR Iris Challenge evaluation in non-cooperative environments: Segmentation and localization," in *Proc. IEEE Int. Joint Conf. Biometrics*, Aug. 2021, pp. 1–10.