

Seg-Cat: Automatic Pupil Detection from Healthy and Cataract Patients

Arth Mishra¹, Naksh Aggarwal², Satkrit³, Sritha Uppaluru⁴,
Vivaan Rungta⁵, Vivaan Rustagi⁶

Abstract

Cataract is among the leading causes of vision impairment worldwide, affecting approximately 94 million individuals who often lack access to timely diagnostic services. Early detection is critical to preventing irreversible vision loss. This paper presents Seg-Cat, a pupil segmentation pipeline designed to address a key limitation of current State-of-the-Art (SOTA) models: the tendency to focus on irrelevant visual features such as skin colour and eyebrow texture rather than the medically relevant pupil region. Seg-Cat employs classical image processing techniques alongside deep learning segmentation using a U-Net architecture, with novel shape-aware loss functions to enforce anatomically consistent pupil boundaries. Experimental results demonstrate that incorporating geometric regularisers (circular and elliptical loss) into the U-Net training objective yields significant improvements over the baseline, achieving a Test Dice score of 0.931 and IoU of 0.871 with the circular variant.

Keywords: Cataract Detection, Pupil Segmentation, U-Net, Circular Loss, Elliptical Loss, Hough Circle Detection, Medical Image Segmentation

1. Introduction

A cataract occurs when the natural lens of the eye becomes progressively cloudy, causing vision to appear blurry, dim, or distorted. Although cataracts are most prevalent in older adults, they can also develop as a result of prolonged ultraviolet (UV) exposure, diabetes mellitus, or direct eye injuries. Globally, nearly 94 million individuals experience vision loss due to cataracts, with a substantial proportion lacking access to appropriate eye care or diagnostic services [1]. Since cataracts develop slowly, patients frequently do not notice symptomatic changes until vision has already been seriously compromised. This underscores the importance of early, automated detection systems that are both accessible and reliable. Automated cataract grading from eye images offers a cost-effective and scalable alternative to specialist examination, particularly in rural or under-resourced settings. A critical challenge with existing automated approaches is that SOTA models tend to focus on spurious visual features such as skin tone, eyebrow density, or scleral discoloration rather than on the pupil, where the cataract manifests optically. This paper introduces Seg-Cat, a framework that first isolates the pupil region from full eye images and then applies a trained segmentation model to produce masks suitable for downstream cataract classification. Four modelling approaches were investigated: a classical Hough-circle based method, a baseline U-Net, a U-Net augmented with a circular loss regulariser, and a U-Net augmented with an elliptical loss regulariser.

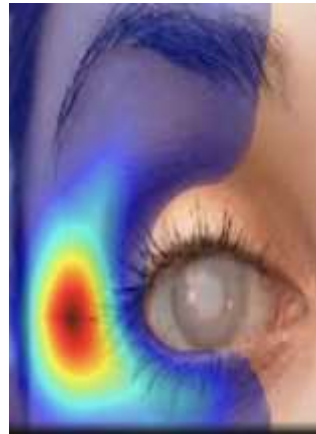


Image 1.1: Model picking up on secondary or unnecessary features

2. Related Works

Convolutional neural networks (CNNs) have been extensively applied to medical image analysis. Yang et al. (2021) demonstrated the effectiveness of CNNs for object detection and semantic segmentation in medical imaging [2]. Albuquerque et al. (2025) provided a systematic review of deep learning-based object detection algorithms in the medical domain, highlighting rapid advances in model accuracy and generalisability [3]. Li et al. (2019) proposed CLU-CNNs for medical image object detection, achieving strong performance on clinical datasets [4]. In the domain of image segmentation, Patil et al. (2013) reviewed classical and modern algorithms for medical image segmentation, establishing foundational frameworks still referenced today [5]. Ramesh et al. (2021) provided a comprehensive survey of segmentation algorithms relevant to pervasive health applications [6]. Olabbarriaga et al. (2001) surveyed human-computer interaction in medical image segmentation, highlighting the role of user guidance in improving segmentation quality [7]. Despite this body of work, there remains a gap in pupil-specific segmentation pipelines designed explicitly for cataract grading. Existing cataract detection models frequently operate on full eye images without isolating the diagnostically relevant lens region, leading to models that rely on confounding features. Seg-Cat addresses this gap by introducing a dedicated pupil segmentation stage as a precursor to cataract classification.

3. Dataset

The dataset used in this study uses an available dataset representing various cataract stages and scales. Cataract severity was classified into six categories following the nuclear sclerosis (NS) grading scale: No Cataract, NS1, NS2, NS3, NS4, and NS5. The distribution of images across classes is as follows: No Cataract (103), NS1 (64), NS2 (659), NS3 (524), NS4 (109), NS5 (180). The NS1 category represents mild, early-stage cataracts, while NS5 represents highly advanced, severe cases.



Image 1.2: The 6 severity grades of cataract.

All images were sourced from mobile camera captures, reflecting real-world clinical acquisition conditions. Ground-truth segmentation masks were generated to indicate the pupil region for supervised training.

4. Methods and Approach

Four distinct methods were investigated for pupil segmentation and cataract detection. These are referred to as Model 1 (Hough Loss / Classical Pipeline), Model 2 (Base U-Net), Model 3 (Circular Loss U-Net), and Model 4 (Elliptical Loss U-Net).

4.1. Overall Framework

The Seg-Cat framework operates in two stages. In the first stage, the pupil region is localised from the full eye image, producing a binary segmentation mask. In the second stage, the masked pupil region is extracted and passed to a cataract classification model. This design ensures that downstream classification is performed exclusively on the medically relevant area, removing confounding background features such as the sclera, surrounding skin, and eyebrows.

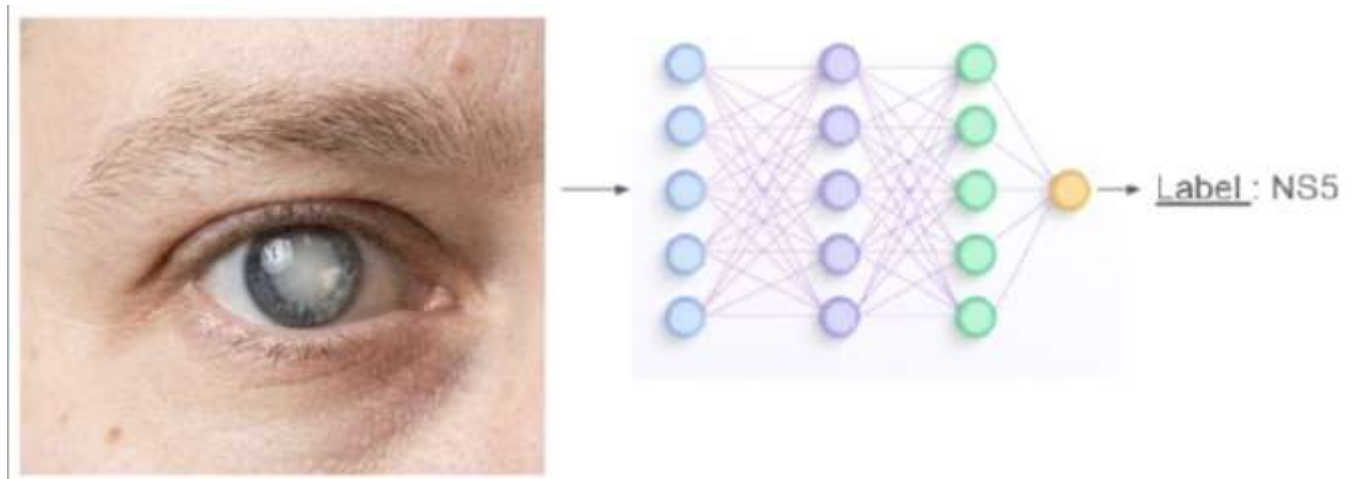


Image 1.3: Seg-Cat’s pipeline, featuring a carefully curated neural network.

4.2. Model 1: Classical Hough Circle Detection

The first approach applies classical computer vision techniques without deep learning. The pipeline proceeds as follows: the input image is converted to the HSV colour space, where thresholding isolates dark pupil-like regions. Morphological operations are then applied to clean noise and fill gaps in the detected region. Finally, Hough Circle Detection [8] is used to identify the circular boundary of the pupil. The resulting circular region is cropped and resized for downstream use. This method serves as an interpretable baseline that does not require labelled mask data for training.

4.3. Model 2: Baseline U-Net

The baseline deep learning model is a U-Net architecture [9] implemented in PyTorch using the MONAI framework. Input images are preprocessed (resized, normalised) before being passed to the network. The model is trained using a hybrid Dice-Focal loss:

$$L = 0.7 \cdot L_{Dice} + 0.3 \cdot L_{Focal}$$

The Dice loss measures overlap between predicted and ground-truth masks, by using the formula:

$$L_{Dice} = 1 - \frac{2|Y_{Pred} \cap Y_{True}|}{|Y_{Pred}| + |Y_{True}|}$$

Focal Loss (with focusing parameter $\gamma = 2.0$) down-weights easy-to-classify pixels to address the severe class imbalance caused by the small pupil region relative to the full image. Starting from binary cross-entropy, $CE(p, y) = -y \cdot \log(p) - (1 - y) \cdot \log(1 - p)$ where $p \in [0, 1]$ is the predicted probability of the positive class and $y \in \{0, 1\}$ is the ground-truth label, Focal Loss multiplies each term by a modulating factor $(1 - p_t)^\gamma$:

$$L_{Focal} = -\alpha \cdot y \cdot (1 - p)^\gamma \cdot \log(p) - (1 - \alpha)(1 - y) \cdot p^\gamma \cdot \log(1 - p)$$

The factor $(1 - p_t)^\gamma$ approaches 0 for well-classified examples and approaches 1 for misclassified ones, so gradient flows mostly through the hard pixels.

Training used cosine annealing learning rate scheduling and mixed-precision arithmetic. Dataset splitting was performed in a stratified manner to preserve class proportions across train, validation, and test sets.

4.4. Model 3: Circular Loss U-Net

Model 3 extends the baseline by introducing a Circularity Regulariser as an additional loss term. The combined loss is:

$$L = 0.6 \cdot L_{Dice} + 0.3 \cdot L_{Focal} + 0.1 \cdot L_{Circ}$$

The circular loss is computed from the predicted boundary points. Let $\{(x_i, y_i)\}_{i=1}^N$ denote the coordinates of the i -th boundary point and N denote the total number of boundary points. The centroid of the predicted boundary is first estimated from the average coordinates of all boundary points $x = \frac{1}{N} \sum_{i=1}^N x_i$ and $y = \frac{1}{N} \sum_{i=1}^N y_i$. Using this centroid, the radial distance of each boundary point from the centre is computed $d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} + \epsilon$. The mean radius is then obtained as the average of all radial distances $\hat{r} = \frac{1}{N} \sum_{i=1}^N d_i$. The circular loss is subsequently defined as the variance of the radial distances around the mean radius $L_{Circ} = \frac{1}{N} \sum_{i=1}^N (d_i - \hat{r})^2 = \text{Var}(d_1, \dots, d_N)$. This loss is minimised when all boundary points are located at approximately the same distance from the centroid, resulting in a boundary that closely resembles a circle. By penalising variations in radial distance, the circularity regulariser encourages geometrically consistent pupil boundaries while preserving segmentation performance and introducing no additional trainable parameters.

4.5. Model 4: Elliptical Loss U-Net

Model 4 replaces the circular regulariser with an elliptical loss term, motivated by two observations. First, the pupil in two-dimensional projection is naturally elliptical rather than perfectly circular. Second, data augmentation (including rotations and affine transforms) introduces additional elliptical distortions that a strict circularity constraint cannot accommodate. The combined loss is:

$$L = 0.6 \cdot L_{Dice} + 0.3 \cdot L_{Focal} + 0.1 \cdot L_{Ellip}$$

The elliptical loss is computed using Principal Component Analysis (PCA) of the predicted boundary points. The centroid of the predicted boundary is first estimated from the average coordinates of all boundary points $c_x = \frac{1}{N} \sum_{i=1}^N x_i$ and $c_y = \frac{1}{N} \sum_{i=1}^N y_i$. Using this centroid, the boundary coordinates

are centred and a covariance matrix is constructed $p_i = [x_i - c_x, y_i - c_y]^T$, $\Sigma = \frac{1}{N} \sum_{i=1}^N p_i p_i^T$ to capture the spatial distribution of the boundary pixels. PCA is then applied to the covariance matrix, where the largest and smallest eigenvalues determine the lengths of the semi-major and semi-minor axes $a^2 =$

$\lambda_{max}(\Sigma)$, $a^2 = \lambda_{min}(\Sigma)$ of the fitted ellipse, respectively. For each boundary point, an algebraic residual is computed $r_i = \frac{(x_i - c_x)^2}{a^2} + \frac{(y_i - c_y)^2}{b^2} - 1$ to measure its deviation from the estimated ellipse. The elliptical loss is subsequently defined as the mean squared residual over all boundary points:

$L_{Ellip} = \frac{1}{N} \sum_{i=1}^N r_i^2$. This loss is minimised when all boundary points satisfy the fitted ellipse equation, allowing the regulariser to adapt naturally to variations in pupil shape and orientation introduced by imaging conditions and data augmentation without requiring additional trainable parameters.

5. Results

Table 1 reports performance metrics for all three deep learning models on the held-out test set. All models were evaluated on Val Dice, Test Dice, IoU, Precision, and Recall.

Model	Val Dice	Test Dice	IoU	Precision	Recall
Hough Transform	0.045	0.035	0.034	0.929	0.062
Baseline UNet	0.913	0.904	0.825	0.891	0.919
Circ + UNet	0.942	0.931	0.871	0.928	0.936
Ellipse + UNet	0.935	0.924	0.859	0.916	0.933

Table 1: Comparison of Model Performance on Test Set

The Circular U-Net (Model 3) achieved the best performance across all metrics, with a Test Dice of 0.931 and IoU of 0.871, representing improvements of 2.7 and 4.6 percentage points respectively over the baseline. The Elliptical U-Net (Model 4) also outperformed the baseline on all metrics, achieving a Test Dice of 0.924 and IoU of 0.859, though it slightly underperformed relative to the circular variant. This suggests that while elliptical geometry is theoretically better suited to the projected pupil shape, the added flexibility may have introduced optimisation challenges with the available dataset size. Visual inspection of predicted masks confirms that both geometry-constrained models produce smoother, rounder boundaries compared to the baseline, which occasionally generates irregular mask edges. The classical Hough-circle pipeline (Model 1) is extremely inaccurate and has only a 11% chance of even detecting a circle.

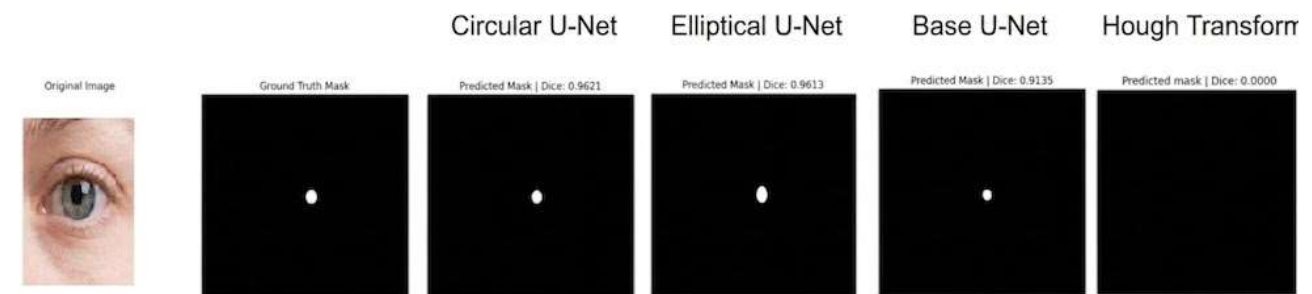


Image 1.4: Visualisation of an image utilizing all four models compared to the ground truth

6. Limitations and Future Work

The primary limitation of the best-performing models is computational cost. Deeper U-Net architectures

combined with shape-based loss computations require significant GPU memory and training time, limiting deployment on low-resource or real-time systems. Although the higher-accuracy models produced better segmentation results, they required **approximately 19.3 times more memory** than the lighter models. This defeated the purpose of designing a small, fast, and accessible system for use by villagers in resource-constrained environments. Model performance may also be sensitive to image quality, lighting conditions, and the accuracy of ground-truth masks. Future work should explore lighter architectures such as Mobile U-Net or reduced-channel U-Net variants to improve inference speed. Model compression and pruning techniques could further reduce computational requirements. The robustness of the pipeline should also be evaluated on images captured under varying lighting conditions and with different mobile devices to assess generalisability beyond the current clinical dataset. Integration of the pupil segmentation stage with a complete end-to-end cataract grading classifier represents the next step towards a deployable diagnostic tool.

7. Conclusion

This paper presented Seg-Cat, a pupil segmentation framework addressing the tendency of existing cataract detection models to focus on non-diagnostic image features. By isolating the pupil region through classical image processing or deep learning segmentation, Seg-CatSe provides a focused input to downstream classifiers. Incorporating geometry-aware loss functions (circular and elliptical regularisers) into U-Net training consistently improved segmentation accuracy over the standard Dice-Focal baseline. The Circular U-Net achieved the best results, demonstrating that anatomical priors can be effectively encoded as differentiable loss terms.

Acknowledgement

The authors thank the rural hospital in Purulia, West Bengal, India for helping us understand the disease etiology of cataract and its manifestation in the images. The team contributions were as follows.

Name	Introduction	Literature Survey	Methodology	Experimentation	Paper Writing
Arth	•	•	•	•	•
Naksh	•	•	•	•	•
Satkrit	•	•	No Contribution	No Contribution	•
Sritha	•	•	•	•	•
Vivaan Run	•	•	•	•	•
Vivaan Rus	•	•	No Contribution	No Contribution	No Contribution

We also thank Cheenta Academy for providing a platform for us to conduct our research work.

References

1. World Health Organization. "Blindness and Vision Impairment." <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
2. Yang R., Yu Y., "Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis", Frontiers in Oncology, 2021, 11, 638182.

3. Albuquerque C., Henriques R., Castelli M., "Deep Learning-Based Object Detection Algorithms in Medical Imaging: Systematic Review", Heliyon, 2025, 11 (1).
4. Li Z., et al., "CLU-CNNs: Object Detection for Medical Images", Neurocomputing, 2019, 350, 53-59.
5. Patil D.D., Deore S.G., "Medical Image Segmentation: A Review", International Journal of Computer Science and Mobile Computing, 2013, 2 (1), 22-27.
6. Ramesh K.K.D., et al., "A Review of Medical Image Segmentation Algorithms", EAI Endorsed Transactions on Pervasive Health and Technology, 2021, 7 (27).
7. Olabarriaga S.D., Smeulders A.W.M., "Interaction in the Segmentation of Medical Images: A Survey", Medical Image Analysis, 2001, 5 (2), 127-142.
8. Duda R.O., Hart P.E., "Use of the Hough Transformation to Detect Lines and Curves in Pictures", Communications of the ACM, 1972, 15 (1), 11-15.
9. Ronneberger O., Fischer P., Brox T., "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical Image Computing and Computer-Assisted Intervention, 2015, 9351, 234-241.