# Towards Precision in Ocular Surface Temperature Measurement: Improvements in the ThermOcular System's Image Segmentation

**Navid Shahsavari**[1,2] **Paul Murphy**[1] **Alexander Wong**[2] **Ehsan Zare Bidaki**[1,2]
[1]School of Optometry and Vision Science, University of Waterloo
[2]Vision and Image Processing Group, System Design Engineering, University of Waterloo
{navid.shahsavari,paul.murphy,alexander.wong,ehsanzb}@uwaterloo.ca

## Abstract

Ocular Surface Temperature (OST) assessment is an important biomarker for ocular health evaluation. Infrared (IR) thermography, prized for its precision and non-invasiveness, stands out in this domain. The ThermOcular system, is a tool for assessing OST and combines machine vision and thermal imaging for corneal segmentation to locate the region of interest. We expanded the ThermOcular's capabilities using state-of-the-art semantic segmentation models to improve and extend the identification of additional ocular components, specifically the pupil, iris, and sclera. Drawing from two datasets, our exclusive 'Apricot' and the renowned TEyeD, and utilizing OCRNet in conjunction with transfer learning, significant enhancements in segmentation performance were achieved . As a result, the upgraded ThermOcular system presents unparalleled precision in OST assessments, paving the way for innovative advancements in ocular health diagnostics.

## 1 Introduction

The ThermOcular system is a sophisticated tool designed for the precise assessment of Ocular Surface Temperature (OST), a crucial metric in ocular health diagnostics[1]. Integrated into this system is a dual-camera setup: one dedicated to infrared(IR) imaging and the other to visual imaging. Positioned on a slit lamp bio microscope base (Figure 1), it assures optimal capture conditions for both imaging modalities. Upon initialization, the system records a synchronized video feed encompassing both the IR and visual spectral ranges of an examined eye. Subsequently, control points are defined manually within these frames, serving as pivotal reference nodes for the image registration algorithm. The importance of this registration cannot be understated, as it ensures spatial congruence between visual and IR frames.

*Fig. 1:* The ThermOcular device used for Ocular Surface Temperature assessment.[1]

Following successful registration, a specialized segmentation algorithm is invoked. Its primary task: the delineation of the Region of Interest (ROI), specifically, the cornea. Through this algorithm, temperature data of the cornea are extracted, ensuring a targeted approach to OST.

To maintain a consistent and automated data extraction process, the system employs an optical flow algorithm. This algorithm autonomously identifies control points for upcoming frame pairs, facilitating a continual stream of temperature data. It's important to note, however, that the primary focus of this paper is not on the optical flow technique itself, but rather on improving the segmentation performace of the system. The intricate steps of this algorithmic procedure are detailed in Figure 2 .

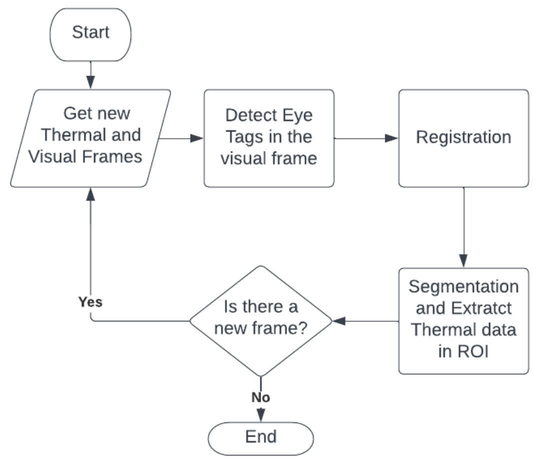


*Fig. 2:* Operational Flowchart of the ThermOcular System

The segmentation process within the ThermOcular device serves a dual purpose: ensuring the specificity of temperature readings by focusing solely on the cornea, and eliminating extraneous IR data from other ocular components. Such precision is vital, not merely for accuracy but also for the reliability of the OST assessments derived from the system.

The ThermOcular system employed the Resnet50-PSPNet[2] as its primary neural network for segmenting the cornea. This model, which is pre-trained on the ImageNet[3], has been further refined using transfer learning technique on a dataset of 126 labeled images specific to the cornea. The effectiveness of this model, gauged using the mean Intersection Over Union (IOU) metric, showcased a commendable performance with an mean IOU of 94.6% for the cornea and background classes.

In this study, OCRNet[4] has been selected for the semantic segmentation task, employing the HRNet-w18 backbone that has been pretrained on the VOC12 dataset[5]. This choice was driven by the need for a more expansive and precise segmentation strategy. The OCRNet was trained on a diverse dataset: 180 labeled images procured from the 'Apricot' study, an initiative undertaken at the Murphy Laboratory of Experimental Optometry (MLEO), coupled with an additional 2,000 images picked from the TEyeD dataset[6]. This comprehensive training bore fruit, with the mean IOU for the pupil, iris, sclera, and background reaching a remarkable 97.61%.

For a meaningful and transparent comparison between the original ThermOcular segmentation model and the one proposed in this study, it's pivotal to focus on equivalent classes. When considering the mean IOU specifically for the pupil and iris (representative of the cornea in the original model) along with the background (rest of the image including sclera), the metric peaks at an impressive 97.61%.

This imporvement corroborates a significant enhancement in segmentation performance, offering a more precise tool for ocular health assessment.

The remainder of this paper is organized as follows: In Section II, the pertinent related works that have shaped the current landscape of ocular segmentation are reviewed. Section III delves into the process of dataset preparation, elucidating the steps undertaken to ensure its variance. Subsequently, in Section IV, a comprehensive comparison of state-of-the-art models trained on the prepared dataset is presented, underscoring each model's respective merits and limitations. In conclusion, the findings and potential avenues for future research are discussed.

## 2 Related Works

Ocular biometry has experienced significant advances in recent years, primarily focusing on iris recognition and segmentation techniques. This section offers insights into pivotal contributions in this realm:

1. **Traditional Iris Segmentation Techniques**:
    - A widely-accepted method that utilized the combination of canny edge detection with circular Hough transform-based algorithms was discussed in [7]. This technique delineates both inner and outer boundaries of the iris, albeit with heightened computational demands.
    - An innovative approach was presented in [8] that employed a sliding rectangular window for initial pupil location. Subsequently, the outer iris circle was ascertained through the assessment of grayscale values outside the pupil zone.
    - The *IrisSeg* framework, introduced in [9], adopted a coarse-to-fine adaptive filtering technique to detect the iris boundary, exemplifying the adaptability of segmentation methods.
2. **Deep Learning in Iris Segmentation**:
    - The domain has witnessed the profound impact of deep learning methodologies. Foremost examples comprise the use of Fully Convolutional Networks (FCN) as seen in [10–12], and the amalgamation of Generative Adversarial Networks (GAN) with FCN in [11]. Such architectures prioritize hierarchical feature extraction.
    - The *PixlSegNet* framework, elucidated in [13], stands out with its convolutional encoder–decoder architecture. The unique integration of a stacked hourglass network between the encoder and decoder paths provided a nuanced segmentation approach.
    - The inclusion of the capsule network in iris recognition, detailed in [14], introduced a dynamic routing algorithm within the capsule layers, which was a deviation from traditional CNN frameworks.
3. **Segmentation Challenges and Advancements**:
    - The research in [15] proposed a distinct methodology rooted in the AdaBoost algorithm and neural networks. Bypassing the common assumption of iris circularity, this method classified iris image pixels without such constraints, enabling greater segmentation adaptability.
    - The studies in [16, 17] pioneered the use of modular neural nets (MNN) for iris detection, underlining the complexities of iris detection in cluttered scenes and the advantages of neural nets in such contexts.
4. **Pupil Segmentation**:
    - The recent inclination towards deep learning finds its application in pupil segmentation as well. *DeepVOG*, as characterized in [18], adopted the U-Net based CNN structure. This model's emphasis on video-oculography (VOG) images demonstrated its high precision, as reflected by the Dice coefficient.
    - A staged methodology was put forth in [19, 20], where the preliminary CNN provides a rudimentary pupil position estimate, which was subsequently refined by a secondary CNN. This tiered strategy aspired to enhance real-time segmentation accuracy.
    - An avant-garde approach in [21] advocated the employment of a deconvolutional neural network for pupil segmentation, tested on an extensive array of datasets.

The evolution from conventional methodologies to contemporary deep learning architectures in ocular segmentation is evident from the aforementioned studies. This work builds upon these foundational concepts, by producing further refinement in segmentation techniques.

## 3 Methodology

### 3.1 Dataset Preparation

To acquire a diverse and representative set of ocular images, a study titled 'Apricot' at the MLEO lab, School of Optometry, University of Waterloo was conducted. Machine vision videos spanning 15 seconds, were recorded for 20 subjects as they blinked naturally. From these recordings, a total of 180 images were carefully selected to encompass a range of eye conditions: fully open, half-open, and completely closed, among others.

Annotation of these images was executed using the Supervisely toolbox, a reliable tool for segmentation-focused image annotations. To further bolster the dataset's robustness, an additional 2000 images from the TEyeD dataset[6], a publicly available dataset were incorporated. (Figure 3)

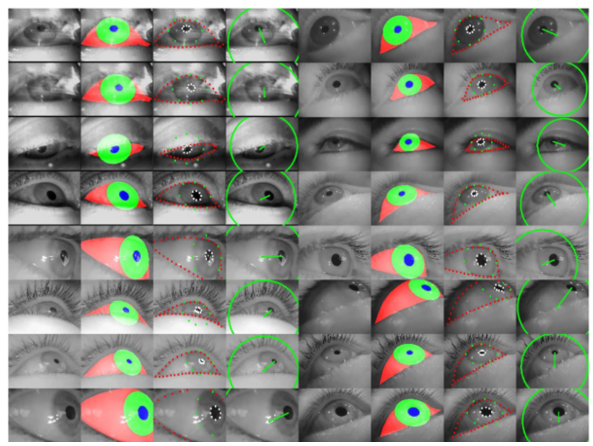

*Fig. 3:* Sample images from TEyeD dataset.

Given the imbalance between the proprietary 'Apricot' images and the TEyeD images, the former were replicated tenfold before augmentation. This was essential to ensure adequate representation during model training. Following this, a series of data augmentations—including rotation, flipping, and cropping—were applied, aiming to enhance the dataset's diversity and consequently improve model generalization.

### 3.2 Segmentation

#### 3.2.1 Model Selection

For ensuring the finest segmentation outcomes, a series of state-of-the-art semantic segmentation models were assessed, including PSPNet[2], Deeplabv3[22], PP-LiteSeg[23], Unet+++[24], SegFormer[25], and OCRNet[4]. These models were selected as they have demonstrated effectiveness in various semantic segmentation challenges, especially in medical images segmentation tasks.

#### 3.2.2 Training Details

The optimal model configuration entailed a batch size of 5 across 15,000 iterations, with a learning rate of 0.01. The SGD optimizer was employed, coupled with the cross entropy loss function. Notably, the HRNet-w18 served as the backbone, with the model being pretrained on the VOC12 dataset, subsequently fine-tuned on the MLEO proprietary dataset.

### 3.2.3 Benchmarking

To gauge the segmentation efficacy, the mean IOU metric was employed. This metric captures the model's accuracy in distinguishing between various ocular regions, offering a holistic view of segmentation performance. The performance of each model, in terms of their overall mean IOU, is tabulated below:

*Table 1:* Benchmarking Results for Semantic Segmentation Models

| Model-Backbone | Pre-train DS | MIOU(total) (%) | MIOU Cornea&bg |
|---|---|---|---|
| PSPNet-ResNet50 | ImageNet | 91.68 | 94.6 |
| Deeplabv3-ResNet50 | VOC12 | 95.28 | 97.11 |
| PP-LiteSeg-STDC2 | Cityscapes | 92.13 | 93.54 |
| Unet | Cityscapes | 90.21 | 93.91 |
| Unet+++ | Cityscapes | 96.34 | 96.88 |
| SegFormer | Cityscapes | 96.34 | 97.36 |
| OCRNet-HRNet-w18 | VOC12 | **97.61** | **98.24** |

## 4 Conclusion

Accurate assessment of Ocular Surface Temperature (OST) is pivotal for thorough ocular health diagnostics. The ThermOcular system, utilizing IR thermography, marks a significant stride in this direction. This paper reports on a method to improve its performance, with a focus on refining the semantic segmentation of ocular components.

The experiments with state-of-the-art semantic segmentation models, notably the OCRNet with an HRNet-w18 backbone, showcased notable improvements, achieving a mean IOU of 98.24% for critical ocular components when compared to the previous model.

These enhancements raise the bar for OST monitoring and set a precedent for future endeavors in ocular imaging and diagnostics. The refined ThermOcular system, with its heightened precision and comprehensive approach, stands as a promising tool in the realm of ocular health assessment, paving the way for further advancements in the field.

## 5 Acknowledgement

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

[1] E. Z. Bidaki, A. Wong, and P. Murphy, "A novel computational thermal-visual imaging system for automatic cornea temperature measurement and tracking," *Journal of Computational Vision and Imaging Systems*, vol. 8, no. 1, pp. 20–23, 2022.

[2] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[4] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 2020, pp. 173–190.

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[6] W. Fuhl, G. Kasneci, and E. Kasneci, "Teyed: Over 20 million real-world eye images with pupil, eyelid, and iris 2d and 3d segmentations, 2d and 3d landmarks, 3d eyeball, gaze vector, and eye movement types," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 367–375.

[7] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.

[8] R. H. Abiyev and K. Altunkaya, "Neural network based biometric personal identification with fast iris segmentation," *International Journal of Control, Automation and Systems*, vol. 7, pp. 17–23, 2009.

[9] 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 *2016 international conference on biometrics (ICB)*. IEEE, 2016, pp. 1–8.

[10] M. Arsalan, R. A. Naqvi, D. S. Kim, P. H. Nguyen, M. Owais, and K. R. Park, "Irisdensenet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors," *Sensors*, vol. 18, no. 5, p. 1501, 2018.

[11] C. S. Bezerra, R. Laroca, D. R. Lucio, E. Severo, L. F. Oliveira, A. S. Britto, and D. Menotti, "Robust iris segmentation based on fully convolutional networks and generative adversarial networks," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2018, pp. 281–288.

[12] Y. Yang, P. Shen, and C. Chen, "A robust iris segmentation using fully convolutional network with dilated convolutions," in *2018 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2018, pp. 9–16.

[13] R. R. Jha, G. Jaswal, D. Gupta, S. Saini, and A. Nigam, "Pixisegnet: pixel-level iris segmentation network using convolutional encoder–decoder with stacked hourglass bottleneck," *IET biometrics*, vol. 9, no. 1, pp. 11–24, 2020.

[14] T. Zhao, Y. Liu, G. Huo, and X. Zhu, "A deep learning iris recognition method based on capsule network architecture," *IEEE Access*, vol. 7, pp. 49 691–49 701, 2019.

[15] Z. B. Zhang, S. L. Ma, P. Zuo, and J. Ma, "Fast iris detection and localization algorithm based on adaboost algorithm and neural networks," in *2005 International Conference on Neural Networks and Brain*, vol. 2. IEEE, 2005, pp. 1085–1088.

[16] H. El-Bakry, "Fast iris detection using neural nets," in *Canadian Conference on Electrical and Computer Engineering 2001. Conference Proceedings (Cat. No. 01TH8555)*, vol. 2. IEEE, 2001, pp. 1409–1414.

[17] H. M. El-Bakry, "Human iris detection using fast cooperative modular neural nets and image decomposition," *Machine Graphics & Vision International Journal*, vol. 11, no. 4, pp. 499–512, 2002.

[18] Y.-H. Yiu, M. Aboulatta, T. Raiser, L. Ophey, V. L. Flanagin, P. Zu Eulenburg, and S.-A. Ahmadi, "Deepvog: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning," *Journal of neuroscience methods*, vol. 324, p. 108307, 2019.

[19] W. Fuhl, T. Santini, G. Kasneci, and E. Kasneci, "Pupilnet: Convolutional neural networks for robust pupil detection," *arXiv preprint arXiv:1601.04902*, 2016.

[20] W. Fuhl, T. Santini, G. Kasneci, W. Rosenstiel, and E. Kasneci, "Pupilnet v2. 0: Convolutional neural networks for cpu based real time robust pupil detection," *arXiv preprint arXiv:1711.00112*, 2017.

[21] F. J. Vera-Olmos and N. Malpica, "Deconvolutional neural network for pupil detection in real-world environments," in *Biomedical Applications Based on Natural and Artificial Computing: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2017, Corunna, Spain, June 19-23, 2017, Proceedings, Part II*. Springer, 2017, pp. 223–231.

[22] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[23] J. Peng, Y. Liu, S. Tang, Y. Hao, L. Chu, G. Chen, Z. Wu, Z. Chen, Z. Yu, Y. Du *et al.*, "Pp-liteseg: A superior real-time semantic segmentation model," *arXiv preprint arXiv:2204.02681*, 2022.

[24] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer, 2018, pp. 3–11.

[25] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 077–12 090, 2021.