RITnet: Real-time Semantic Segmentation of the Eye for Gaze Tracking

Aayush K.Chaudhary*[1], Rakshit Kothari*[1], Manoj Acharya*[1], Shusil Dangi[1], Nitinraj Nair[2], Reynold Bailey[2], Christopher Kanan[1], Gabriel Diaz[1], Jeff B. Pelz[1]

[1]Chester F. Carlson Center for Imaging Science , [2]Golisano College of Computing and Information Sciences
Rochester Institute of Technology

RIT College of Science Chester F. Carlson Center for Imaging Science

facebook research OpenEDS Challenge

PerForM Lab

CGAP Computer Graphics & Applied Perception Laboratory

KLab

Multidisciplinary Vision Research Laboratory

Overview

- Accurate eye segmentation improves **eye-gaze estimation**
- Current segmentation models suffer from **person-dependent accuracy**, lack of **robustness** and inability to run in **real-time**.
- We present ***RITnet***, a deep neural network architecture, which obtains **state-of-the-art** results on the 2019 OpenEDS[1] Semantic Segmentation Challenge.
- Model Size: **<1 MB** to allow **real-time** gaze tracking

Architecture

- ***RITnet*** (248,900 trainable parameters) is based on DenseUNet-K[2] architecture which combines U-Net[3] and DenseNet[4] architectures.
- Direct '**skip**' connections with simplified **dense** connected feature maps
- Constant channel size of 32 to reduce model complexity
- The back-propagated loss is:

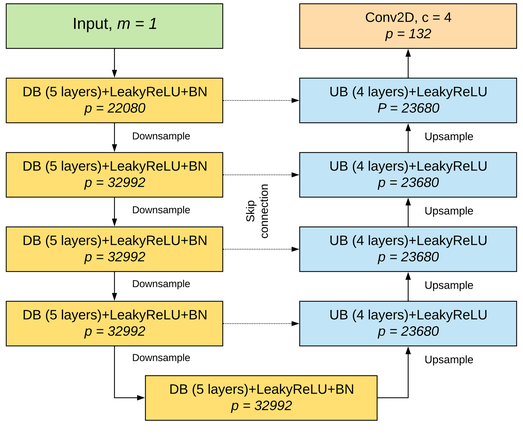
$$L = L_{CEL}(\lambda_1 + \lambda_2 L_{BAL}) + \lambda_3 L_{GDL} + \lambda_4 L_{SL}$$

Cross Entropy Loss (L_{CEL})

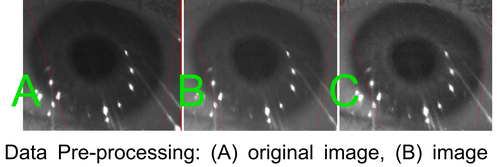
Generalized Dice Loss (L_{GDL})

Boundary Aware Loss (L_{BAL})

Surface Loss (L_{SL})



Architecture details of RITnet



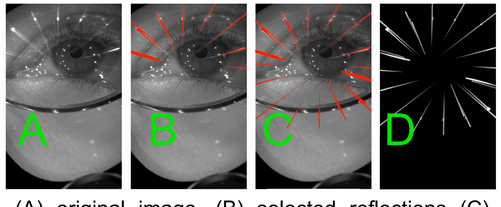
Data Pre-processing: (A) original image, (B) image after Gamma correction with a factor of 0.8, and (C) image after local Contrast Limited Adaptive Histogram Equalization

Experimental Details

- OpenEDS Semantic Segmentation dataset consists of *train (8916)*, *validation (2403)* and *test (1440)* images.
- Each image contains four semantic labels: *pupil, iris, sclera*, & *background*.
- Adam optimizer with learning rate 0.001 and batch size of 8 images for 175 epochs.
- **Loss scheduling scheme** to prioritize production of **crisp boundary pixels**, handle **class imbalance** and penalize **stray patches**.

Augmented the training images by

- Horizontal flip
- Translation (0-20 pixels)
- Gaussian Blur ((7,7), $2 \leq \sigma \leq 7$)
- Thin lines around random center
- *Starburst* pattern from train image



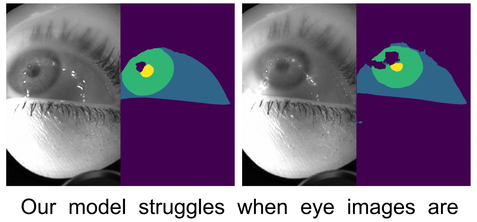
(A) original image, (B) selected reflections (C) concatenating with 180° rotation (D) final mask which is appended with training image.

Results

- Evaluation is based on mean Intersection over Union (mIoU) and the number of model parameters (s) as

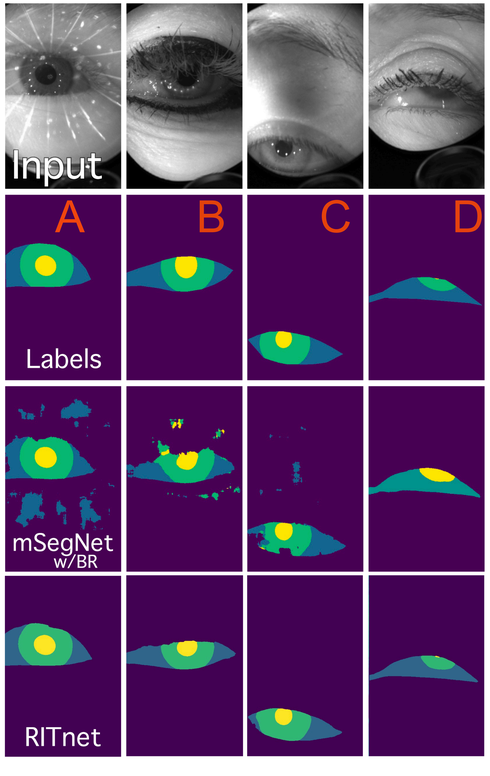
$$(mIoU + min(1/s, 1))/2$$

- Our model achieves a ~**6%** improvement in mIoU score while the complexity is reduced by ~**38%** compared to the baseline model modified SegNet with Separable convolutions (SC)[1]



Our model struggles when eye images are **heavily blurred** or **defocused.**

Model	Mean F1	mIoU	Model Size (S)	No. of parameters (million)	Overall Score
mSegNet*	97.9	90.7	13.3	3.5	0.491
mSegNet* w/BR	98.3	91.4	13.3	3.5	0.495
mSegNet* w/SC	97.4	89.5	1.6	0.4	0.762
Ours	**99.3**	**95.3**	**0.98**	**0.25**	**0.976**



Comparison of results in the challenging conditions (A) multiple reflections due to prescription glasses, (B) heavy mascara (C) dim light (D) partial eyelid closure.

Contributions

- State-of-the-art results of 95.3% with model size of <1 MB
- Real-time segmentation at >300 Hz for 640x400 images on an NVIDIA 1080Ti GPU
- Domain-specific augmentation schemes
- Boundary aware loss functions with a loss scheduling strategy

References

1. S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi. OpenEDS: Open Eye Dataset. 4 2019.
2. S. Dangi and C. Linte. DenseUNet-K: A simplified Densely Connected Fully Convolutional Network for Image-to-Image Translation. https://github.com/ShusilDangi/DenseUNet- K/blob/master/DenseUNet K.pdf, 9 2019.
3. O. Ronneberger, P. Fischer, and T. Brox. U-net: Con- volutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.*, 9351:234–241, 2015.
4. G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2017-Janua, pages 2261–2269. IEEE, 7 2017.