

Weakly-supervised Medical Image Segmentation with Gaze Annotations: Appendix

Yuan Zhong¹, Chenhui Tang¹, Yumeng Yang², Ruoxi Qi², Kang Zhou¹,
Yuqi Gong¹, Pheng Ann Heng¹, Janet H. Hsiao³(✉), and Qi Dou¹(✉)

¹ Dept. of Computer Science and Engineering, The Chinese University of Hong Kong

² Dept. of Psychology, The University of Hong Kong

³ Division of Social Science, The Hong Kong University of Science and Technology

A. Gaze Data Collection

Table 1. Eye tracking settings to build the GazeMedSeg dataset using SR Research Experiment Builder. Major settings for the program building are reported.

Eye Tracker	SR Research EyeLink 1000 Plus
Tracked Eye	Monocular (ocular dominance)
Sampling Rate	1000 Hz
Tracking Error	$\leq 0.5^\circ$ visual angle
Screen-Eye Distance	46 – 55 cm
Screen Size	21.5 inch
Screen Resolution	1024 \times 768
Screen Refresh Rate	60 Hz
Displayed Image Size	768 \times 768
Displayed Image Position	Centered
Displayed Image Extent	Vertical & Horizontal: ≈ 54 cm
Using Chin Rest	Yes
Wearing Glasses	No (contact lenses are allowed)
Tracking Program	SR Research Experiment Builder
Host PC OS	Microsoft Windows 10

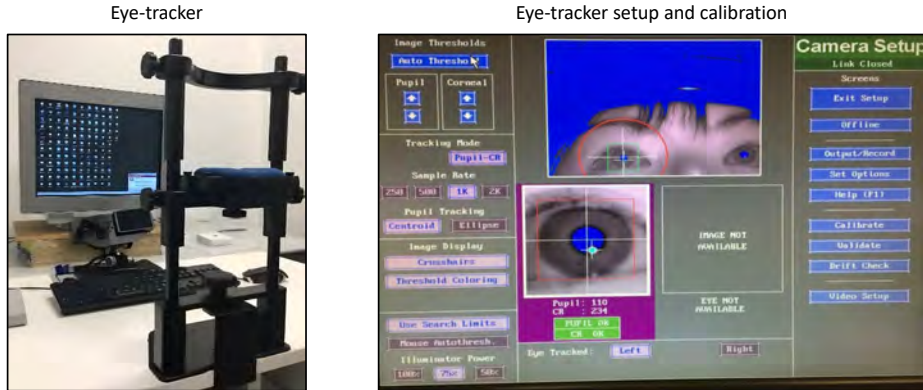


Fig. 1. Eye-tracking equipment and setup. We use a chin rest to stabilize the eye-tracking process. During gaze collection, only one image is displayed on the screen. Annotators use their eyes to annotate and their hands to press a key to switch to the next image once they have finished.

B. Implementation Details

Table 2. Hyper-parameters used in experiments on two datasets.

Backbone	2D UNet
Supervision Loss	Cross-entropy loss
Training Iterations	15000
Batch Size	8
Optimizer	SGD
SGD Momentum μ	0.99
Scheduler	CosineAnnealingLR
Base Learning Rate	$1e^{-2}$
Minimum Learning Rate	$1e^{-4}$
Resolution	224×224
Data Augmentation	Random Flip
Data Split	Kvasir-SEG: 900 training and 100 testing images NCI-ISBI: 789 training and 117 testing images