Using Convolutional Neural Networks with External Webcams to Develop a Reliable and Economical Webcam-Based Eye Tracker

1. Introduction

Neurodegenerative diseases (ND) such as Alzheimer's Disease and dementia are increasing rapidly as the population ages. Over the past two decades, the percentage of individuals 65 years of age or older (yo) in the US has increased by > 54% (8). Cases for Alzheimer's Disease, for example, have been doubling every five years for patients above 65 yo (7). As the number of individuals diagnosed with NDs increases, the urgency to develop more reliable and earlier diagnostics tools is also becoming more prevalent (29).

Many times, subtle symptoms of NDs are exhibited at earlier stages of disease progression where treatment is proven to be more effective. An early marker of NDs include irregular ocular motions (4, 5, 6). Specifically, irregular saccades (rapid eye movements directed towards specific objects) are strongly correlated with NDs; specifically those affected by Alzheimer's Disease and other forms of dementia have slower saccades, less accurate saccades, and longer saccade latencies (4, 5, 6).

Since many forms of dementia, including Alzheimer's, affect ocular motions, eye tracking technologies have recently emerged in an effort to diagnose NDs and other cognitive disorders earlier and have proved promising in research so far (4, 5, 6). Traditional Eye Trackers (ETs) are high technology external machines that measure and capture a participant's gaze point (the location where a person is looking on the computer screen mapped as an x.y coordinate). Using infrared lighting to accurately discern different parts of the eye (including the iris, cornea, and pupil), ETs capture in high resolution at high frequency (frames per second or FPS)

recording the direction in which infrared light is reflected off the eye (15). Research has been conducted utilizing ETs to designate some diseases such as Alzheimer's Disease and eye diseases such as diabetic retinopathy, glaucoma, and macular degeneration, however they are not yet used for any professional diagnoses since such research is preliminary (4, 5, 6, 15).

While the cognitive impairment piece of ETs are a very important application, ETs are versatile and can be used in a variety of other scenarios. A paralyzed individual or those with mobility impairments can use ETs as a computer mouse (16). For example, the individual could move the cursor with their eyes, and blink a determined amount of times in quick succession to indicate a left or a right click. Furthermore, ETs can be used to construct heat maps to identify which parts of a stimulus stick out to certain people.

The aforementioned research is preliminary and further research in this field is inhibited by the high cost of ETs. Similarly, many of the other applications discussed regarding an ET-mouse and heat maps are not feasible with the expensive nature of ETs. Tobii® is a standard giant in the eye tracking industry, with the majority of research done as of yet conducted using Tobii's products. Tobii's regular ETs cost around $8,900, and high end ETs, which are generally required for measuring microsaccades relevant to NDs research, **cost as much as $39,900**. Furthermore, the experimental research software for these ETs, Tobii Pro Lab, costs an additional $8,500 (20).

Therefore, despite ETs holding promise in several fields, accessibility is limited. Convolutional neural networks have been proposed as a method for creating an ET through using a regular webcam (22, 23, 24). While some webcam-based eye trackers are available online, their average errors are considerably worse than that of Tobii's, the industry standard (15, 20, 22, 23, 24). This research uses another software-based approach with special consideration for effective

calibration, image processing, and model architecture. The main goal is to train a model that mimics a traditional ET (such as the Tobii) and achieves a competitive accuracy. By diagnosing NDs early in development, NDs can be controlled more effectively since disease progression could be slowed (9).

2. Experimental Design

Recording Specifications. The research's stimulus was displayed on a 13 inch laptop. The laptop's webcam was used to capture images. The webcam captured at 30 FPS in 720p. A water bottle was placed 60 cm in front of the computer to act as a chin rest. The participant rested their chin on the water bottle to prevent excessive face movement.

Participants. For this research, the student researcher (myself) was the only participant.

Capturing Programs. All of the project's related program components were programmed in Python. Each sitting, the program familiarized itself with the recording setting and the person first by building a training set. The participant rested their head back against the headrest of the chair, and followed a dot that moved across the monitor with their eyes. Three webcams captured the eyes repeatedly as the dot moved and associated the image with a label including the location of the dot at each instance. About 750 images and their corresponding locations were captured for each webcam in this program. 600 of these instances were captured during the Training Set Capture (Fig. 1), and 150 instances were captured in the Validation Set Capture (Fig. 2), which ran immediately subsequent to the Training Set Capture.

Figure 1. Training Set Capture (calibration). Program Illustration of the paths the dot followed for the Training Set Capture, the first of the stimuli to be presented. The webcam captured continuously as the dot moved across the screen, capturing 600 images over the duration. When the dot reset its path on a new horizontal line, the dot and camera paused for two seconds in order to allow the participant to readjust their eyes to the new position. This program took about two minutes to complete.

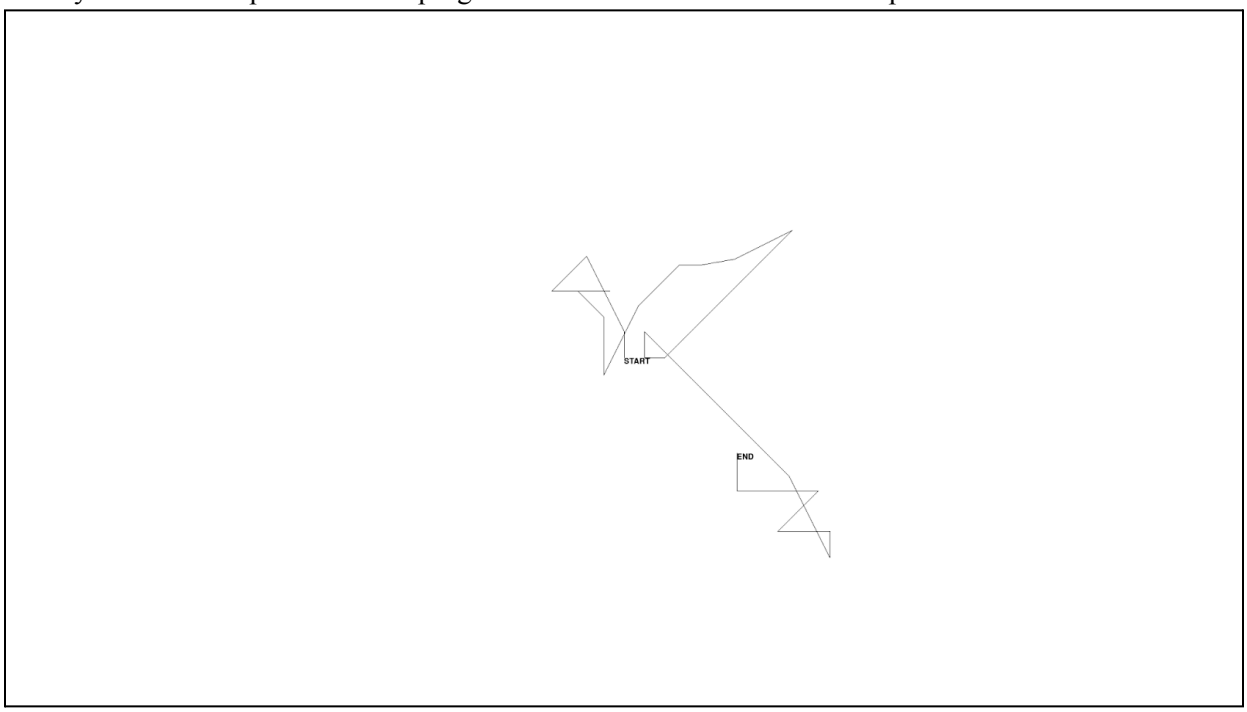

Figure 2. Validation Set Capture. Program Illustration of an iteration of a path the dot followed for the Validation Set Capture. The dot's movement was random, moving in eight possible different directions at

different speeds for different durations. The above illustration shows one instance of the random behavior, and this behavior changed every time the program ran. The webcam captured continuously as the dot moved across the screen, capturing 150 images over the duration and took under one minute to complete. This program also ordered the paths the dot followed in order to construct the visualization. In other words, it stored the order of the paths that the dot followed for visual analysis since the movement was random. The dot, which always started in the center of the screen, had its beginning of the path demarcated with the printed text in the image reading "START", and its last position demarcated with "END". Images captured from this run were compared to the true dot path to validate the model's efficacy, and a similar illustration to the one above was built using only the eye images captured during the run of this program.

Image Processing. All 750 images captured (including images from both sets) went under image processing. The webcam captured images of not just the eyes but the entire face and some of the body from the shoulders up and the background, so the image was cropped through a facial recognition library that detected where the eyes were in the entire image (14). Following cropping, the image was converted into black and white through a set thresholding method. This threshold varied for each data set, and the value of the threshold was fine tuned to best provide accurate black and white eye images for the recording setting. This threshold does not need to be adjusted after done once for the recording setting. After conversion, the white pixels that represented the sclera and the black pixels that represented the pupil were enumerated. If either of these values were not large enough (as determined on a percentile basis), it was assumed there was no sclera/pupil in the eye image, and so the image was removed from the image set. After image processing, the image set will have eliminated eye images that would have raised issues with learning and led to large errors. An example image set from the Training Set Capture (Fig. 1) is shown in Figure 3.

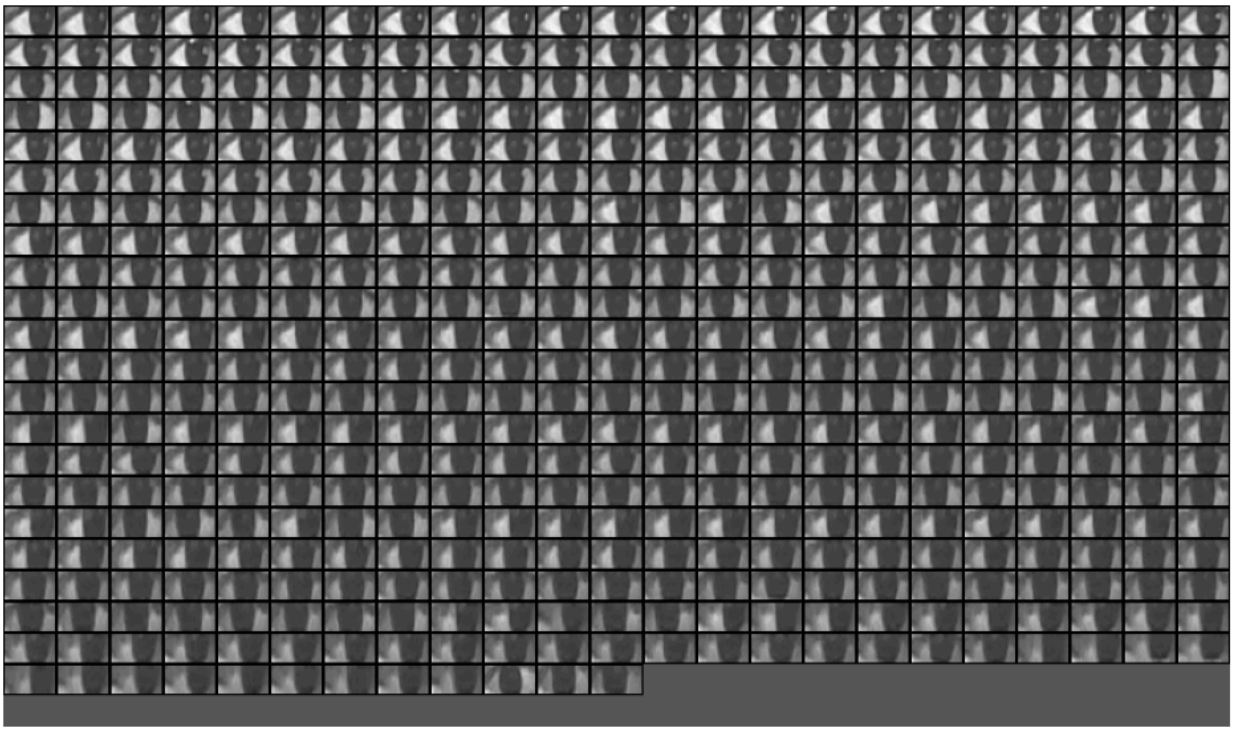

Figure 3. *Example of Training Set for Left Eyes*. There are 583 images in this image, as 17 images were removed as they were detected as outliers, and therefore removed.

Learning Parameters. The image set and associated labels were split into a training set and a validation set. 95% of the 600 images captured in the first program (the calibration program illustrated in Fig. 1) were used to train the model. The other 5% of these images were reserved to track Validation Loss. The 150 images captured in the second program (the validation program illustrated in Fig. 2) were plugged into the model after training to verify that the model could work on the new, unseen images. The training images were split into batches of 8 for learning. Each set of eyes was trained separately through the resnet50 pre-trained convolutional neural network. The model ran for 600 epochs. Cross Entropy Loss was used for loss since it is an optimal loss function for CNNs. During training, the resnet50 CNN was used with a Stochastic Gradient Descent (SGD) optimizer, a learning rate of 0.001, and a momentum of 0.9. The SGD optimizer was used as the optimizer since it could generalize data it sees so it returns better

values for images not seen before by the model. Loss and Validation Loss was plotted live through the online Weights And Biases machine learning platform to verify progress of the model.

Validation Calculation. The model was trained to predict a location coordinate on the screen at which it thought the person was looking at. Images for validation locations were input to the trained model and the average coordinate output of each eye was compared to the true coordinates. An approximation of the degree error, the standard unit of error in eye tracking, was calculated through a trigonometric tangent calculation. The distance (in mm) between the predicted and true coordinates was the "opposite" side and the 600 mm measured from the participants' face to the monitor was the "adjacent" side. A visualization based off Figure 2 including the predicted locations was built as a visualization for the errors.

3. Results

Both eyes, which were recorded in the same capture, were trained separately in the model and exhibited strong indications of learning throughout the task. Loss graphs and Validation Loss graphs had decreasing trends with decreasing variability for most runs across the epochs (Fig. 4, Fig. 5).

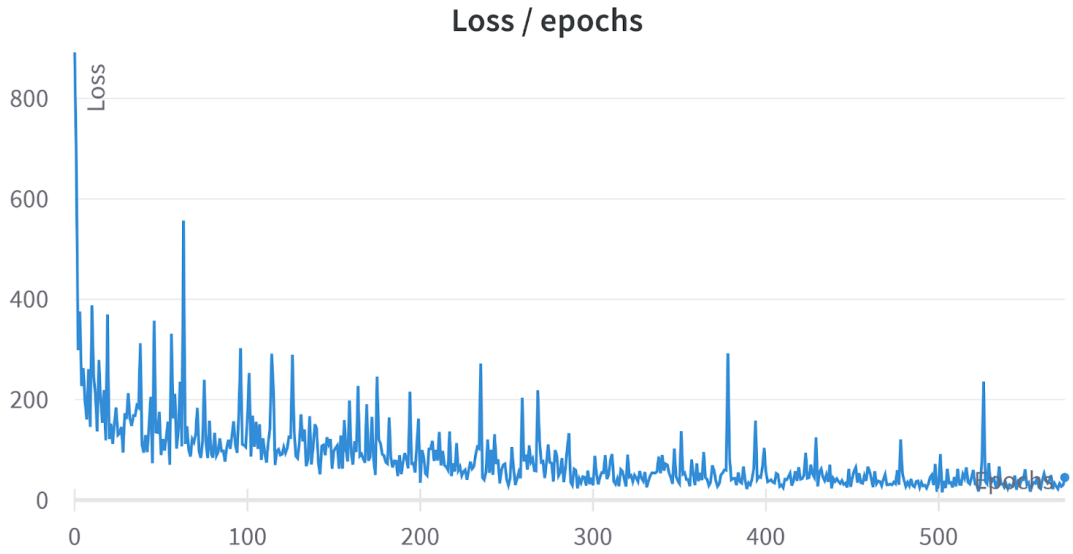

Figure 4. Loss graph of a sample run. The loss graph tracked the relative error of images in the training set (95% of images captured by the calibration program in Figure 1). The decreasing nature of the graph demonstrated that the CNN was able to perform the task and was improving over the epochs. Note: the X-axis is measured in epochs.

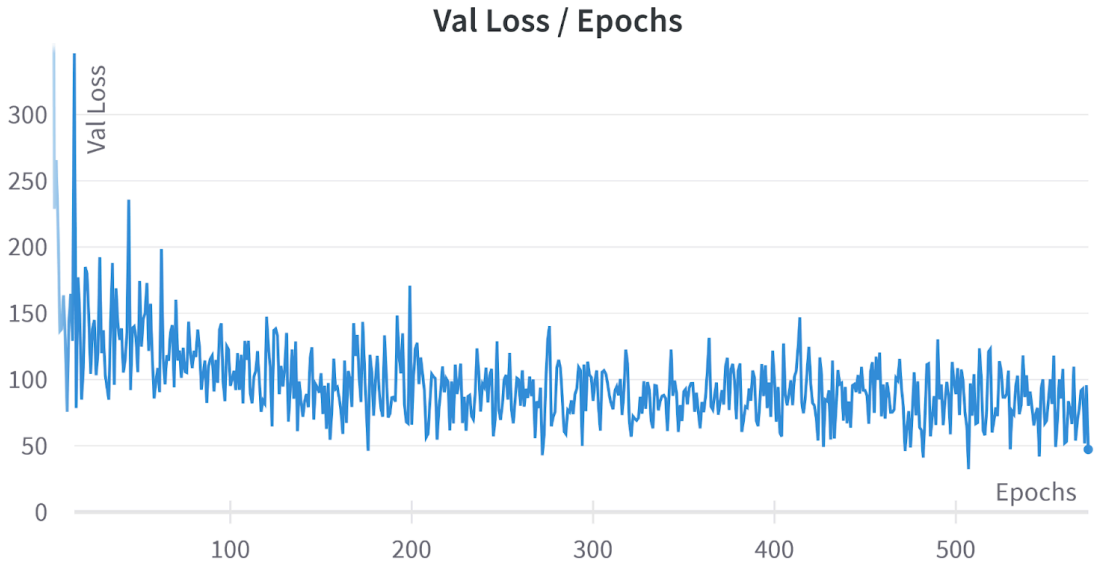

Figure 5. Example Validation Loss Graph. The validation loss graph tracked the relative error of images in the validation set (5% of images captured by the calibration program in Figure 1). The 150 images from the Validation Set Capture were not included in this data set; that set is reserved for constructing a visualization.

Subsequent to learning, images for each specific validation location were passed through

the model and the average coordinate value between the two eyes were assessed. Average errors for this CNN webcam-based eye tracker typically ranged from 0.72° and 1.57° for all validation images. Validation images captured in locations unlike ones within the training set often yielded higher errors. Visualization of errors corresponding to validation images captured in Figure 2 is illustrated (Figures 6 and 7).

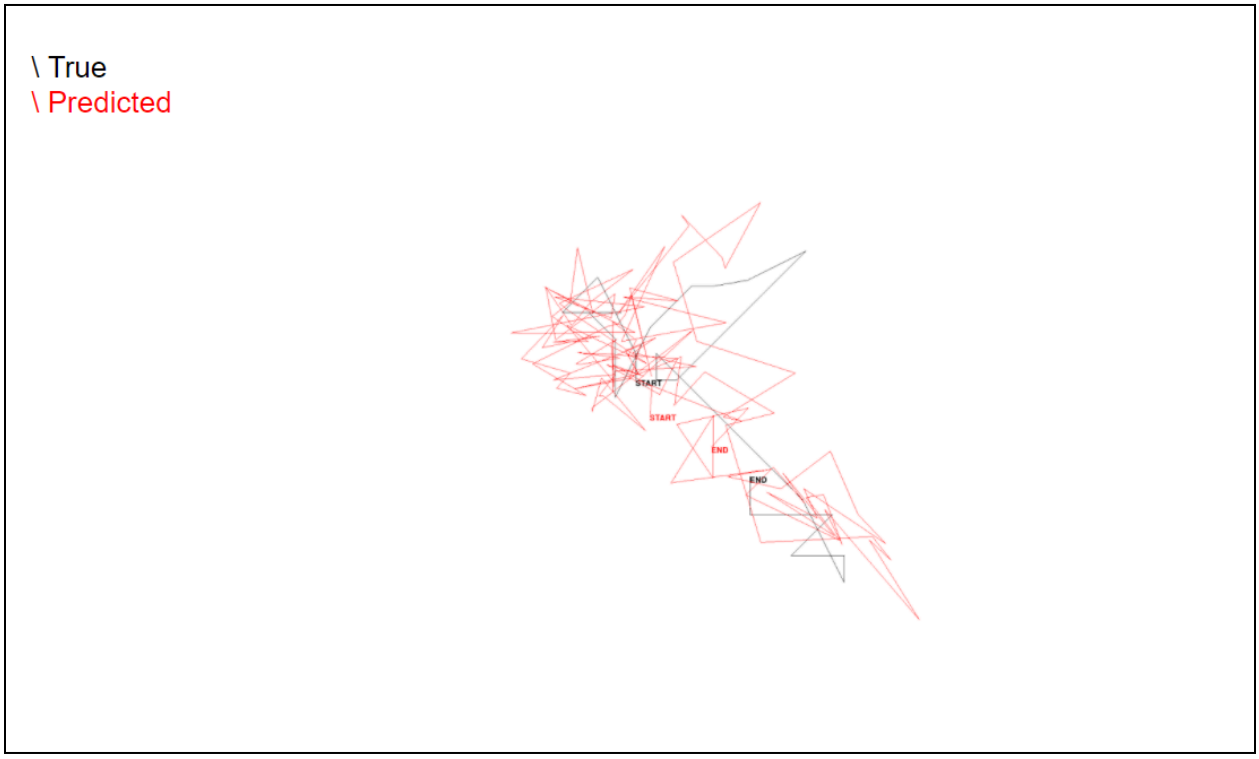

Figure 6. Illustration of Errors by predicting all images captured in the Validation Set Capture (Fig. 2). This was the visualization from the validation set capture which had 150 images minus several outliers (143 images are represented here, to be precise). The red lines represent the predicted path while the black lines represent the dot's true path. The illustration may seem noisy due to outliers that were not detected by the model and hence not removed. These outliers caused errors as great as 2.25° which may make the overall predictions seem off. This program had an average error of 0.74° despite all the noise.

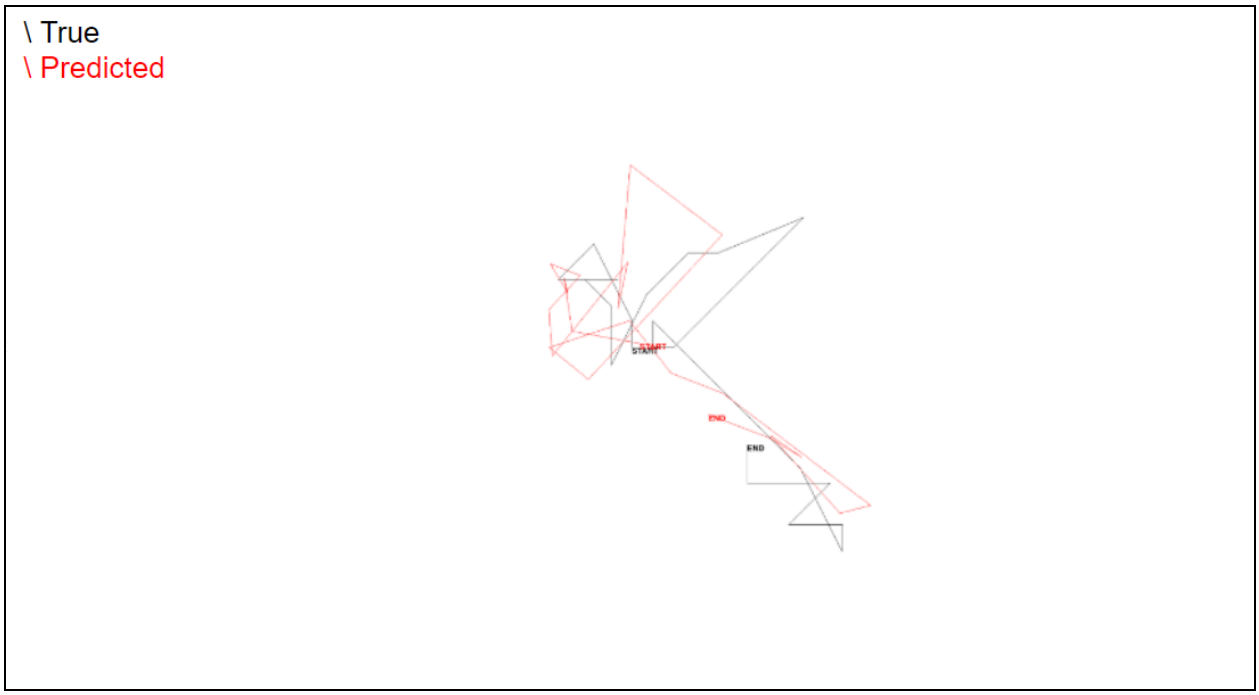

Figure 7. Illustration of Errors, cleaned up. This image is from the same instance of the data set from Figure 6, however includes fewer predictions. In this visualization, only end points (vertices) in the Validation Set Capture were illustrated. Images captured when the dot was following a path were removed in this visualization, leading to fewer red lines. There were only 25 predictions in this program, which was the same number of black vertices in the illustration. See Figure 8 for more details. This trial, after cleaning up, had a very similar average error of about 0.74°.

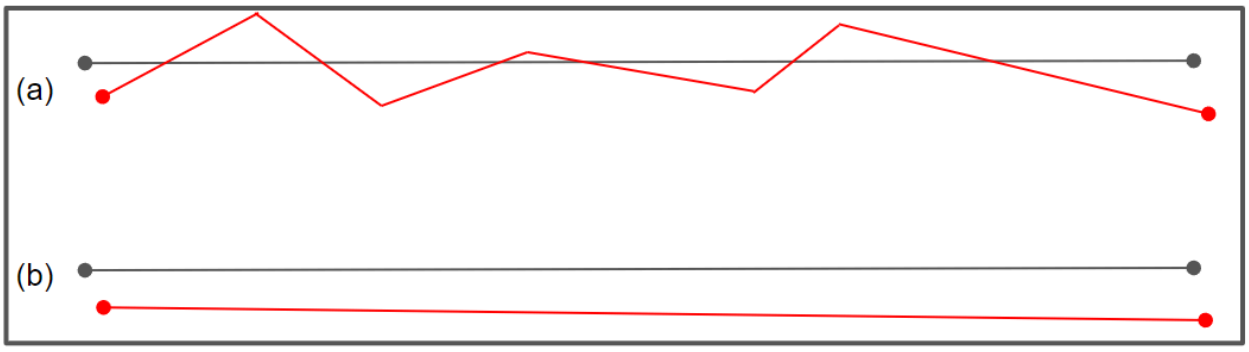

Figure 8. Explanation of Cleaning Up the Illustration of Errors (Fig. 7). In the Validation Set Capture, the program captured images continuously as the dot moved across a line. The previous visualization included predictions (illustrated by the vertices in red) from all images, including intermediate images as the dot moved across the straight line (illustrated in black), leading to noise as summarized in (a). Figure 7 included only end points as predictions, making the visualization less noisy, as summarized in (b).

4. Discussion

Accuracies produced in this research are comparable to existing technologies and show further proof of concept for webcam-based eye tracking. By comparison, The Tobii Pro Spectrum® is an eye tracker costing $39,900, excluding the Tobii Pro Lab experimental software (which costs up to an additional $8,600) (20). The Tobii Pro Spectrum averages errors of roughly 0.4° and 0.58° based on both internal and external validation studies (12, 32).

GazePointer is a webcam-based eye tracking method that was validated only internally. Their internal validation yielded a mean error of 1.05° (24). GazePointer used a $90 external webcam in their study, and the inner workings of their eye tracking system architecture is not described (24).

An immediate future project includes utilizing three external webcams that capture in 1080p at 60 frames per second for this software, and to validate on human participants other than myself (the student researcher) to gain confidence in the results. Three webcams remove bias of face movements and posture changes that affect the viewpoint from which one is looking at (27). It should be noted that this could be accomplished with only two webcams, however additional webcams will likely improve accuracy (27). This removes the need for a chin rest, and functions similar to how the Tobii includes multiple cameras (28). The capturing rate of 60 frames per second allows for basic saccadic measurements such as saccade peak velocities to be calculated (30). External webcams with these specifications that could be used will also be cheap, costing less than $20 each (31). Regardless of the cost, this future project would be also adaptable to the number of webcams available, meaning the option to use a single webcam, as done in this research, will be available and pose no additional cost.

The use of a webcam-based eye tracker in a clinical setting holds significant promise as it achieves the project's larger purpose of improving the early diagnosis of cognitive impairment for more avenues of effective treatment. A phone application, for example, could mimic an eye tracker in the future (which has cameras capable of capturing at 120 FPS, allowing for substantive saccadic calculations) to provide provisional and preliminary insight into cognitive function. Depending on the results of the eye tracking test, it could suggest a person look further into their cognitive health by consulting with a professional.

5. References

1. Anderson, T. J., & MacAskill, M. R. (2013, January 22). Eye movements in patients with neurodegenerative disorders. Nature News. Retrieved August 18, 2021, from https://www.nature.com/articles/nrneurol.2012.273.
2. Leonard F. M. Scinto, P. D. (1994, July 1). Impairment of spatially directed attention in patients with probable Alzheimer's Disease as measured by eye movements. Archives of Neurology. Retrieved August 18, 2021, from https://jamanetwork.com/journals/jamaneurology/article-abstract/592951.
3. Molitor, R. J., Ko, P. C., & Ally, B. A. (2015, January 6). Eye movements in Alzheimer's Disease. Journal of Alzheimer's Disease : JAD. Retrieved August 18, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5332166/.
4. Sun, T. (2021). TeleAEye: Low-Cost Automated Eye Disease Diagnosis Using a Novel Smartphone Fundus Camera With AI. Full Abstract. Retrieved August 18, 2021, from https://abstracts.societyforscience.org/Home/FullAbstract?ISEFYears=0%2C&Categ

ory=Translational+Medical+Science&AllAbstracts=True&FairCountry=Any+Country&FairState=Any+State&ProjectId=21255.

5. Garg, R. (2019). Diagnosing Autism with Machine Learning: Binary Classification for Eye Movement in Virtual Reality Environment. Full Abstract. Retrieved August 18, 2021, from https://abstracts.societyforscience.org/Home/FullAbstract?ProjectId=17363.
6. Chia, T. (2021). A-EYE: Utilizing Multistage Neural Networks and Landmark Localization for Fundus Image Disease Detection. Full Abstract. Retrieved August 18, 2021, from https://abstracts.societyforscience.org/Home/FullAbstract?ISEFYears=0%2C&Category=Translational+Medical+Science&AllAbstracts=True&FairCountry=Any+Country&FairState=Any+State&ProjectId=20919.
7. Liu, Z., Yang, Z., Gu, Y., Liu, H., & Wang, P. (2021, July 12). The effectiveness of eye tracking in the diagnosis of cognitive disorders: A systematic review and meta-analysis. PLOS ONE. Retrieved August 21, 2021, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0254059.
8. Naqvi, E. (2017, June 30). *Alzheimer's Disease Statistics*. Alzheimer's News Today. Retrieved September 29, 2021, from https://alzheimersnewstoday.com/alzheimers-disease-statistics/.
9. Sheinerman, K. S., & Umansky, S. R. (2013, January 1). *Early detection of neurodegenerative diseases: Circulating brain-enriched microrna*. Cell cycle (Georgetown, Tex.). Retrieved August 29, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3570496/.
10. Bureau, U. S. C. (2021, September 26). Census.gov. Retrieved November 29, 2021, from

https://www.census.gov/.

11. Farnsworth, B. (2021, October 11). *Eye tracker prices - an overview of 20+ eye trackers*. Imotions. Retrieved August 12, 2021, from https://imotions.com/blog/eye-tracker-prices/.
12. *Tobii Pro Spectrum product description - NBT*. Tobii Pro. (n.d.). Retrieved February 2, 2022, from https://nbtltd.com/wp-content/uploads/2018/05/tobii-pro-spectrum-product-description.pdf
13. Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018, May 23). *An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood*. Frontiers. Retrieved February 2, 2022, from https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00803/full
14. Ageitgey. (2018, April 2). *Ageitgey/face_recognition: The world's simplest facial recognition API for python and the command line*. GitHub. Retrieved February 2, 2022, from https://github.com/ageitgey/face_recognition
15. Tobii. (2015, August 10). How do Tobii Eye Trackers work? Tobii Pro. Retrieved February 2, 2022, from https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/
16. Wright, T. (2014, December 26). *Eye-tracking technology allows paralyzed people to control pcs, tablets*. Dell Technologies. Retrieved February 2, 2022, from https://www.dell.com/en-us/blog/eye-tracking-technology-allows-paralyzed-people-to-control-pcs-tablets/
17. Ladders Inc. (2018, November 6). *Ladders updates popular recruiter eye-tracking study*

with new key insights on how job seekers can improve their resumes. Ladders Updates Popular Recruiter Eye-Tracking Study With New Key Insights on How Job Seekers Can Improve Their Resumes. Retrieved February 2, 2022, from https://www.prnewswire.com/news-releases/ladders-updates-popular-recruiter-eye-tracking-study-with-new-key-insights-on-how-job-seekers-can-improve-their-resumes-300744217.html

18. Berkovsky, S., Wang, E., Taib, R., & Koprinska, I. (2019, April). *Detecting Personality Traits Using Eye-Tracking Data*. ResearchGate. Retrieved February 2, 2022, from https://www.researchgate.net/publication/332747592_Detecting_Personality_Traits_Using_Eye-Tracking_Data
19. Mikhailenko, M., Maksimenko, N., & Kurushkin, M. (2022, March 10). *Eye-tracking in immersive virtual reality for education: A review of the current progress and applications*. Frontiers. Retrieved April 2, 2022, from https://www.frontiersin.org/articles/10.3389/feduc.2022.697032/full
20. *Contact US - Tobii Pro*. Contact Us - Tobii Pro. (n.d.). Retrieved July 2, 2022, from https://www.tobiipro.com/contact/
21. *Amazon.com: Webcam with microphone, 1080p full HD webcam streaming ...* Webcam with microphone, 1080p full HD webcam streaming. (n.d.). Retrieved July 2, 2022, from https://www.amazon.com/Microphone-Streaming-Computer-Conferencing-Recording/dp/B087WT6L6B
22. YouTube. (2019). *Eye-tracking Mouse Using Convolutional Neural Networks and Webcam*. *YouTube*. Retrieved October 20, 2021, from https://www.youtube.com/watch?v=iV9ZkvdsL7I.

23. Andersen, P. (2018, September 12). *XLabs eye gaze tracking software*. The Wonder of Science. Retrieved February 2, 2022, from https://thewonderofscience.com/phenomenon/2018/7/12/xlabs-eye-gaze-tracking-software
24. Deja, S. (2021, February 5). *Gazepointer: Real-time webcam eye-tracking software*. GazeRecorder. Retrieved February 2, 2022, from https://gazerecorder.com/gazepointer/
25. *IPhone*. Apple. (n.d.). Retrieved February 2, 2022, from https://www.apple.com/iphone/
26. BBC. (2021, December 10). *Brighter, steadier, smarter: How smartphone cameras will improve in 2022 | looking ahead with Tecno | BBC Storyworks*. BBC News. Retrieved February 2, 2022, from https://www.bbc.com/storyworks/future/looking-ahead-with-tecno/brighter-steadier-smarter-how-smartphone-cameras-will-improve-in-2022
27. McCann, S. (n.d.). *3D Reconstruction from Multiple Images*. CS231A - Computer Vision: From 3D reconstruction to recognition. Retrieved September 2, 2022, from https://cvgl.stanford.edu/teaching/cs231a_winter1415/
28. *Does head movement affect eye tracking results?* Improving your research with eye tracking since 2001 - Tobii Pro. (2015, August 17). Retrieved February 2, 2022, from https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/does-head-movements-affect-eye-tracking-results/
29. Gitler, A. D., Dhillon, P., & Shorter, J. (2017, May 1). *Neurodegenerative Disease: Models, mechanisms, and a new hope*. Disease models & mechanisms. Retrieved February 2, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5451177/
30. Wierts, R., Janssen, M. J. A., & Kingma, H. (2008, December). *Measuring saccade peak*

velocity using a low-frequency sampling rate of 50 Hz. IEEE Xplore. Retrieved February 2, 2022, from https://ieeexplore.ieee.org/document/4711478

31. Goettsche Partners. (2011). *60FPS Webcam with Ring Light, Auto-Focus 1080P Web Camera with Dual Microphone and Privacy Cover , Streaming Webcam for YouTube, Skype, Zoom, Twitch, OBS, Xsplit and Video Calling*. Amazon. Retrieved July 16, 2022, from https://www.amazon.com/gp/product/B09PMVZWKJ/ref=ppx_yo_dt_b_search_asin_title?ie=UTF8&psc=1
32. Nyström, M., Niehorster, D. C., Andersson, R., & Hooge, I. (2021, February). *The TOBII Pro Spectrum: A useful tool for studying microsaccades?* Behavior research methods. Retrieved February 16, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7880983/