HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 2 (2024)

Machine Learning Techniques for Biometric Unsupervised Gaze Estimation in ADHD Screening

Yaejoon Jung[1] and Eunju Moon[#]

[1]Phillips Academy Andover, USA
[#]Advisor

ABSTRACT

Attention-Deficit/Hyperactivity Disorder (ADHD) is a complex neurodevelopmental condition characterized by persistent challenges in attention, hyperactivity, and impulsivity, significantly impacting daily functioning and developmental trajectories. The traditional ADHD diagnostic process typically involves comprehensive assessments conducted by healthcare professionals, such as psychologists, psychiatrists, or pediatricians. These assessments rely heavily on subjective observations and reports from various sources. However, this approach is time-consuming, labor-intensive, and often requires multiple appointments which makes it a resource-intensive process. To address this issue, I propose an unsupervised learning-based gaze estimation system for the screening of ADHD. The proposed system takes eye images as input and generates gaze vectors which indicate the individual's current focal point. By aggregating these gaze vectors over a specific time series, the system can identify abnormalities in the gaze patterns which facilitate the early screening of ADHD. Comprehensive experiments have shown the superiority of the proposed system over previous methods. The experimental results also confirm the feasibility of utilizing the proposed method as a biometric for screening ADHD.

Introduction

The conventional diagnostic process for ADHD relies on comprehensive assessments conducted by healthcare professionals, involving subjective observations and reports from multiple sources, such as individuals, parents, and teachers. This traditional approach, while valuable, is fraught with challenges, including its time-consuming and resource-intensive nature, often necessitating multiple appointments. The current diagnostic paradigm poses limitations in terms of efficiency, accessibility, and objectivity. The reliance on subjective assessments may introduce biases and hinder timely intervention. Recognizing these challenges, there is a pressing need for innovative methodologies that can augment or even revolutionize the ADHD diagnostic process.

This research addresses the shortcomings of traditional ADHD diagnosis by proposing a groundbreaking solution – an unsupervised learning-based gaze estimation system for ADHD screening. The core premise of this system involves the analysis of eye images to generate gaze vectors, providing insights into an individual's focal points. Through the aggregation of these gaze vectors over specific time series, the system discerns abnormalities in gaze patterns, offering a novel and objective approach to the early screening of ADHD.

The primary objective of this research is to advance the field of ADHD diagnostics by introducing and evaluating the proposed gaze estimation system. This system, built on unsupervised learning-based gaze estimation, aims to not only streamline the diagnostic process but also contribute to the early identification of ADHD traits.

The structure of this research paper is as follows: Chapter 2 provides background knowledge to enhance the understanding of the proposed method. In Chapter 3, every detail of the proposed method is explained, encompassing its architectural overview and machine learning training procedures. Chapter 4 investigates a comprehensive analysis of experimental results which aim to examine the effectiveness of the proposed method. Finally, Chapter 5 summarizes the key findings and contributions of this research.

Related Work

Gaze Estimation

Gaze Estimation is a modern eye-tracking mechanism utilizing various machine learning methods. In simple terms, this technology allows machines to recognize the direction one is looking at by training the algorithm with certain training samples. As an input mechanism, the algorithm intakes a photo of an eye or a photo encompassing a part of an eye. As its output, it deduces the exact direction of the eye using metrics, yaw and pitch. To elaborate, yaw and pitch respectively are labels that measure the rotation of an object. Yaw measures the vertical angle and pitch measures the horizontal angle of two designated points. In the context of this gaze estimation and its application, yaw and pitch is measured to see how the gaze of a person or eye direction changes over a duration of time.

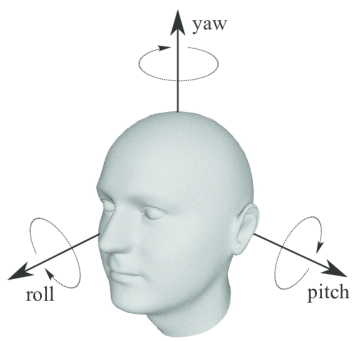

Figure 1. Illustration of yaw, pitch, and roll

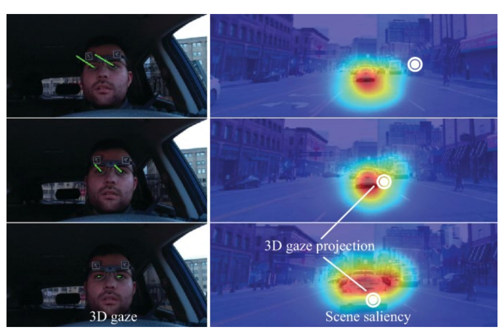

Figure 2. Example of gaze estimation technique (driver monitoring system) (Kasahara et al. 2022)

One of the domains where gaze estimation has found widespread application is in enhancing human-computer interaction. By tracking users' eye movements, systems can intuitively respond to their gaze, enabling hands-free control, cursor navigation, and interactive experiences.

This has implications for virtual reality, gaming, and assistive technologies, making interfaces more natural and accessible. In the automotive industry, gaze estimation contributes to enhancing driver safety (Kar et al. 2017). By monitoring drivers' gaze patterns, systems can detect signs of drowsiness or distraction, issuing warnings or interventions when necessary (Shah et al. 2022). This technology is necessary in the development of advanced driver assistance systems and the progression toward autonomous vehicles. Another application is pilot training simulations (Babu et al. 2019). In terms of training pilots, this technology helps the pilots and pilot instructor to know the objective cognitive ability of a pilot in real-time. Gaze estimation increases training efficiency and training benefit for military aviation fields.

In healthcare, gaze estimation has promising applications, including neurological disorder diagnostics. Gaze-based diagnostics offer a non-invasive and objective method for early detection and monitoring of various health-related issues. In this research, I utilize gaze estimation as a biometric system to screen ADHD. The detailed explanation will be provided in Chapter 3.

Regression in Machine Learning

Machine learning encompasses a diverse set of techniques that enable computers to learn patterns from data and make predictions or decisions. This chapter provides an in-depth exploration of regression, its principles, algorithms, and applications. Regression is a supervised learning technique that focuses on predicting a continuous outcome variable based on one or more predictor variables. In simpler terms, it models the relationship between input features and a numeric target variable. Unlike classification, where the goal is to predict a categorical label, regression aims to estimate a quantity.

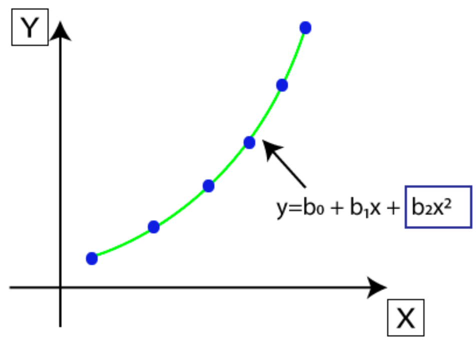

Figure 3. Regression example (fitting polynomial function to data points)

Gaze estimation system can be considered as a regression problem, given that its output, the gaze vector, comprises two continuous values: yaw and pitch (Park et al. 2019). The training of the gaze estimation model involves utilizing loss functions commonly employed in general regression problems, such as mean squared error or mean absolute error. In the proposed approach, I introduce unsupervised learning for gaze estimation to improve the accuracy of the system.

ADHD and Eye Movements

The correlation between ADHD and eye movement has been a subject of research interest aiming to understand the potential relationship between neurodevelopmental disorders and oculomotor behavior. Various studies have investigated how individuals with ADHD may exhibit distinct patterns of eye movement compared to those without the disorder.

Research explains that individuals with ADHD may display differences in saccadic eye movements (Goto et al. 2010). Saccades are rapid, voluntary eye movements that redirect the gaze from one point to another. Some studies propose that individuals with ADHD might exhibit increased variability in the amplitude and velocity of saccadic movements, potentially reflecting underlying cognitive and attentional processes (Levantini et al. 2020). Smooth pursuit eye movements, which involve tracking a moving object smoothly with the eyes, have also been examined in relation to ADHD (Ross et al. 2000). It is hypothesized that individuals with ADHD may experience difficulties in maintaining smooth pursuit, possibly due to challenges in sustaining attention on a moving target.

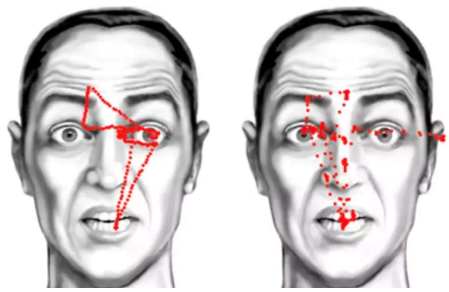

Figure 4. Eye movement for a single image for a five seconds lapse of time. (Galgani et al. 2009) (left) a control subject; (right) a subject diagnosed with ADHD

Motivated by these insights, this research introduces an ADHD screening system that leverages gaze estimation to automatically assess individuals' eye movements by predicting yaw and pitch vectors. Additionally, an unsupervised learning approach is incorporated to augment the system's accuracy. Chapter 3 will explain the details of the proposed system including a comprehensive overview, the system architecture and the choice of loss function employed in training the machine learning model.

Proposed Method

The proposed gaze estimation-based ADHD screening system consists of two integral modules: the gaze estimation module and the ADHD screening module. In the gaze estimation module, an eye image is inputted to generate a gaze vector comprising yaw and pitch values. To enhance the precision of gaze estimation, I introduce a novel unsupervised representation learning method which is explained in detail in Chapter 3.1. The ADHD screening module leverages these gaze vectors to identify anomaly patterns in individuals' eye movements. This process involves predefined post-processing steps, outlined in Chapter 3.2. The flowchart of the proposed system is illustrated in Figure 5.

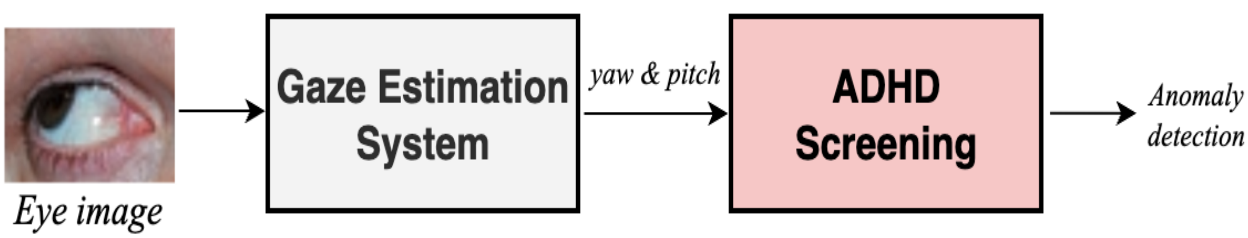

Figure 5. Flowchart of the proposed system

Gaze Estimation

The goal of gaze estimation is to predict the gaze vector which is composed of yaw and pitch values. This system takes eye images as input which include various features such as skin color, brightness, eye shape, and gaze-related features. Notably, among these features, only gaze-related features significantly impact the performance of gaze estimation (Park et al. 2019). It is evident that appearance features like brightness, skin color, or eye shape do not contribute to the efficacy of gaze estimation. Therefore, achieving heightened accuracy in gaze estimation necessitates the disentanglement of gaze-related features from the entangled set of features.

To address this, I propose a novel unsupervised gaze representation learning method that involves applying a rotation matrix to the feature space. This approach aims to enhance the discernment of gaze-related features, thus contributing to the overall accuracy of gaze estimation.

Unsupervised Gaze Representation Learning

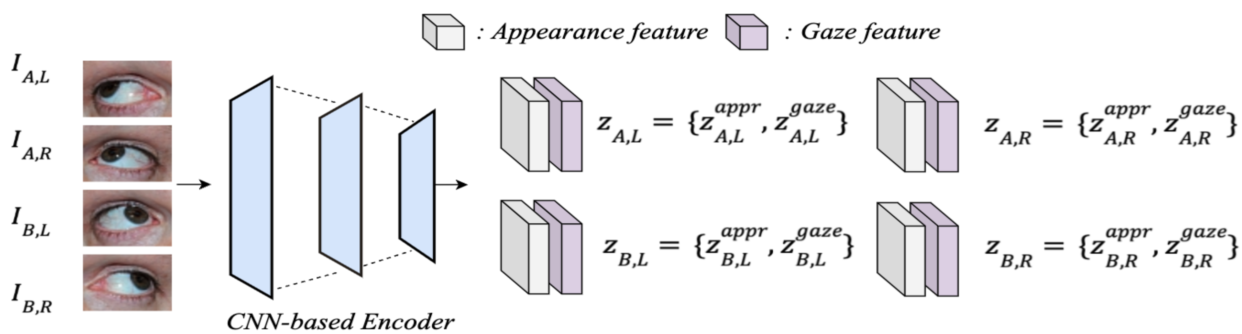

Figure 6. Architecture of the proposed unsupervised gaze representation learning approach (encoder)

The objective of gaze representation learning is to consistently extract and disentangle gaze-related features that significantly impact the enhancement of gaze estimation. To achieve this, I leverage a convolutional neural network-based autoencoder architecture, commonly employed in representation learning. The autoencoder comprises an encoder and a decoder. The encoder processes four pairs of eye images, captured from the same individual but directed towards different gaze points (e.g., $I_{A,L}$ and $I_{A,R}$ for the left and right eye images looking at gaze direction A, and $I_{B,L}$ and $I_{B,R}$ for gaze direction B).

These eye images are inputted to the encoder to generate four feature maps as illustrated in Figure 6. Each eye image is transformed into a feature map, denoted as z, comprising two elements: appearance feature z^{appr} and gaze feature z^{gaze}. As $z_{A,L}$ and $z_{B,L}$ are extracted from the same left eye, they share identical appearance features. Similarly, $z_{A,R}$ and $z_{B,R}$, being derived from the right eye, should exhibit the same appearance features. Additionally, $z_{A,L}$ and $z_{A,R}$ possess matching gaze features, while $z_{B,L}$ and $z_{B,R}$ share equivalent gaze features. The proposed gaze representation learning capitalizes on this inherent consistency by employing a gaze rotation matrix. Each eye image is associated with a ground truth gaze vector, enabling the straightforward derivation of the rotation matrix H between these two gaze vectors, as outlined in Equation 1 (Cuemath 2021).

Equation 1: Calculation of rotation matrix H

$$H^{(\alpha,\beta)} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

Here, $H^{(\alpha,\beta)}$ represents the rotation matrix that transforms vector ***A*** into vector ***B*** and α and β denote the yaw and pitch angle difference of the gaze vector A and B, respectively.

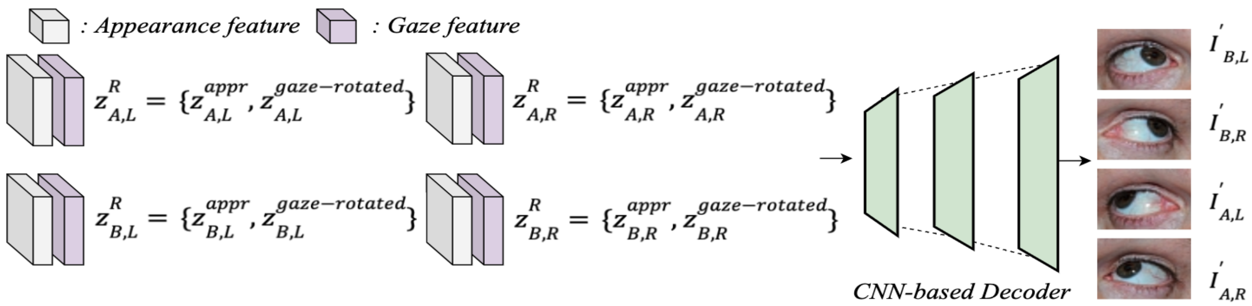

Figure 7. Architecture of the proposed unsupervised gaze representation learning approach (decoder)

The rotation matrix $H^{(\alpha,\beta)}$ is applied in feature space as explained in Equation 2.

Equation 2: Applying rotation matrix in feature space

$$z^{gaze-rotated}_{i,j} = H^{(\alpha,\beta)}\, z^{gaze}_{i,j}$$

Where, $z^{gaze-rotated}_{i,j}$ denotes the rotated gaze-related feature map for i∈{***A***, ***B***}, and j∈{***L***, ***R***}. The computed rotation matrices are applied to all four pairs of feature maps, as illustrated in Figure 7. In the decoding process, these feature space-wise rotated feature maps are utilized to reconstruct the rotated eye images. For instance, when the decoder receives $z^{R}_{A,L}$ as input, it reconstructs $\boldsymbol{I_{B,L}}$ instead of $\boldsymbol{I_{A,L}}$, given that the gaze-related feature is rotated. The fundamental assumption guiding this approach is that the appearance feature remains consistent; however, due to the rotation of the gaze-related feature in feature space, the decoder reconstructs the rotated eye image. To quantify the discrepancy between the generated image and its ground truth, the reconstruction error is measured using the loss function shown in Equation 3.

Equation 3: Reconstruction loss function

$$L = \sum_{i=\{A,B\}} \sum_{j=\{L,R\}} \left| I'_{i,j}(x,y) - I_{i,j}(x,y) \right|$$

Within Equation 3, $I'_{i,j}(x,y)$ denotes the pixel intensity at the specified x and y coordinates of the reconstructed eye image, while $I_{i,j}(x,y)$ corresponds to the pixel intensity at the same coordinates in the ground truth image. Throughout the training phase, the encoder learns to disentangle the gaze-related feature from the appearance feature. The resulting trained encoder exhibits a generalization ability which enhances the accuracy of gaze estimation in subsequent transfer learning. The efficacy of the proposed approach will be thoroughly examined in Chapter 4.

Gaze Estimation

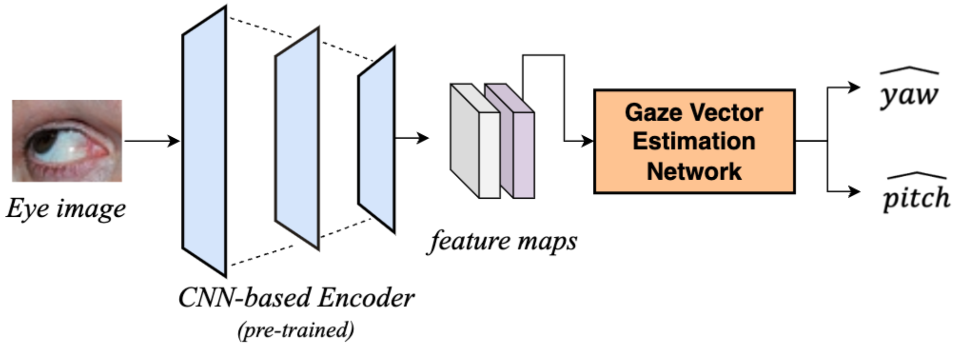

Figure 8. Architecture of the gaze estimation network

Following representation learning, the pre-trained encoder functions as the initial point for fine-tuning the gaze estimation network. The input eye image is fed into the encoder to produce feature maps. From these maps, only the gaze feature map is extracted and inputted into the gaze vector estimation network which predicts gaze vector (yaw and pitch). By adopting this transfer learning strategy, rather than training the network from scratch, the model ensures faster training and achieves accurate results. This contributes to the proposed gaze representation learning. The effectiveness of this approach is investigated in Chapter 4.

For the network architecture, I designed a two-layered neural network for the gaze vector estimation network. The training process employs the mean squared error function which is commonly utilized in regression problems, as shown in Equation 4.

Equation 4: Mean squared error function

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{yaw} - yaw\right)^2 + \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{pitch} - pitch\right)^2$$

In Equation 4, N denotes the total number of samples while $\widehat{yaw}$ and $\widehat{pitch}$ are predicted yaw and pitch of the gaze vector, respectively.

ADHD Screening

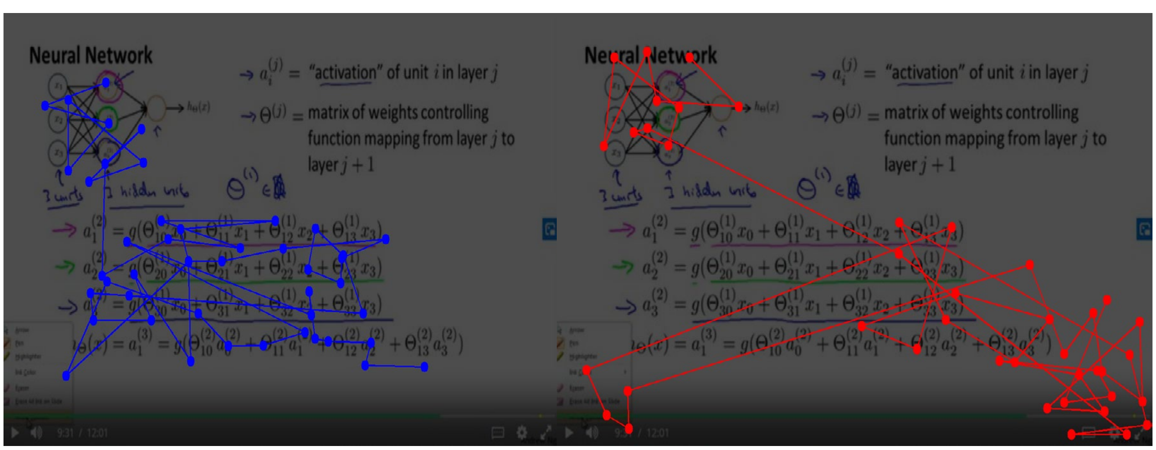

Figure 9. ADHD screening system demonstration. (Left: Gaze vector trajectories of a normal individual; Right: Gaze vector trajectories of an individual with ADHD)

In this chapter, I present the prototype of the ADHD screening system proposed in this research paper. Figure 9 provides a visual representation of the ADHD screening system in action, depicting scenarios for both normal individuals and those with ADHD. Over a specific time period, the system performs gaze estimation, collecting projected gaze vectors on the screen to indicate the individuals' focal points at different moments. Subsequently, the system plots the trajectories formed by these gaze vectors, offering valuable insights for screening purposes to determine the presence of ADHD. Notably, this system is adaptable to any laptop environment equipped with a webcam.

These trajectories can be readily utilized for training a machine learning-based anomaly detection network, representing a key aspect of my future research plans. I expect that this novel approach could significantly enhance the efficiency of early ADHD screening for young students, particularly due to its superior accessibility compared to traditional screening methods for ADHD.

Experimental Results

Dataset

In this chapter, I present an in-depth explanation of the dataset employed in this research study. GazeCapture is a publicly available and open-source dataset specifically curated for training gaze estimation models.

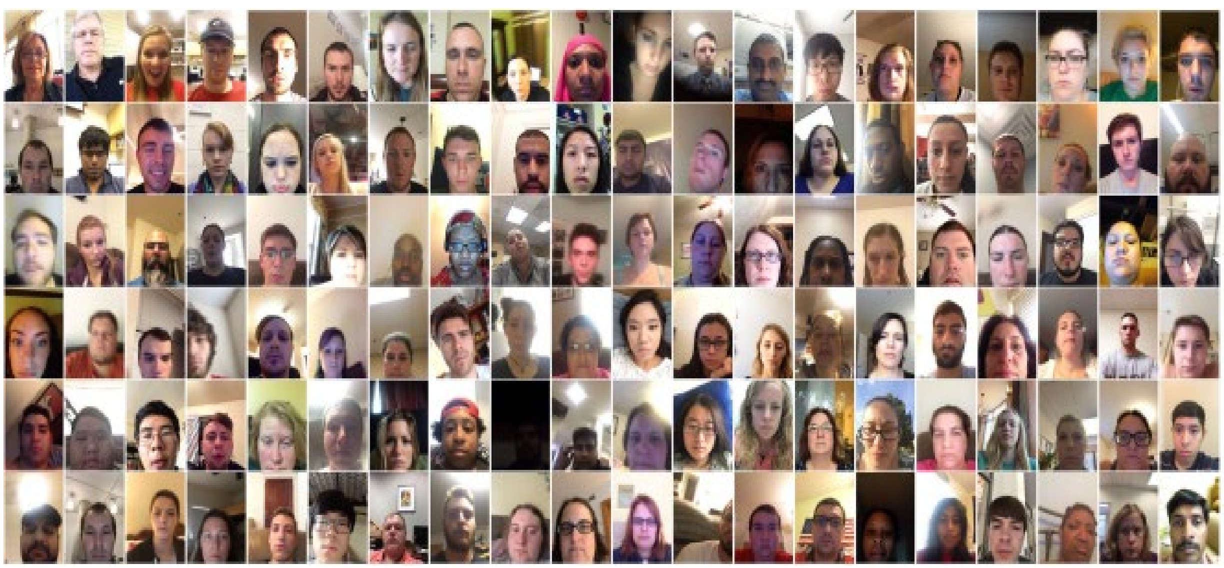

Figure 10. GazeCapture Dataset (Krafka et al. 2016)

GazeCapture contains a diverse and extensive collection of face images sourced from 1,474 individuals. The dataset is crafted to encompass a wide array of facial expressions, head poses, and environmental conditions. The dataset comprises a total of 2,445,504 discrete images. For every image in the dataset, the ground truth gaze vector, represented by the angles (yaw, pitch), is annotated. For the experiments, the dataset is segmented into two subsets: an 80% portion designated as the training set and the remaining 20% allocated as the testing set.

Volume 13 Issue 2 (2024)

Evaluation Metrics

To measure the accuracy of the proposed model, I utilize angular error as an evaluation metric. Angular error is measuring the margin of error of two sets of vectors; in this experiment, it is used to compare the vector proposed by the gaze estimation algorithm and the actual direction of the eye. The formula for angular error is illustrated in equation 5.

Equation 5: Angular Error

$$Angular\ Error = cos^{-1}(cos(\alpha - \beta)) = \frac{A \bullet B}{||A||\,||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Where, *A* and *B* denote the predicted gaze vector and its ground truth, respectively. In the numerator, the dot product of each vector is denoted. In the denominator, the norm of each vector is multiplied together. After a value comes out from this process, this number is inputted to the inverse cosine function. As a result, the equation gives the angle error between the predicted gaze vector and its ground truth.

In the best case scenario when there is no angular error, the equation gives 1 as the value before being processed by the inverse cosine function. In the worst case scenario, this number is -1. As the angular error function is a cosine similarity function, the best and worst case scenarios are described as the two furthest points in the cosine graph.

Experimental Protocol

To compare the performance of the proposed method to the state-of-the-art gaze estimation methods, I conducted two types of experiments. Firstly, the performance of the proposed method was compared with previous state-of-the-art methods within the Gazecapture dataset, referred to as within-dataset evaluation. This evaluation aimed to assess how well the trained approach performs on the specific dataset.

The second type of experiment involved cross-dataset evaluation. In this case, the models were initially trained using the Gazecapture dataset, and then the trained model was tested using a different dataset from a distinct domain. This step was taken to gauge the generalization capabilities of the proposed approach in enhancing gaze estimation performance across diverse datasets.

Evaluation

Within Dataset

Table 1. Evaluation comparison on GazeCapture dataset

Method	Angular Error
FAZE (Park et al. 2019)	13.7
U-LinFT	10.6

(Yu et al. 2020)	
CrossEncoder (Sun et al. 2021)	8.9
MultiGaze (Gideon et al. 2022)	8.0
Proposed Method	6.7

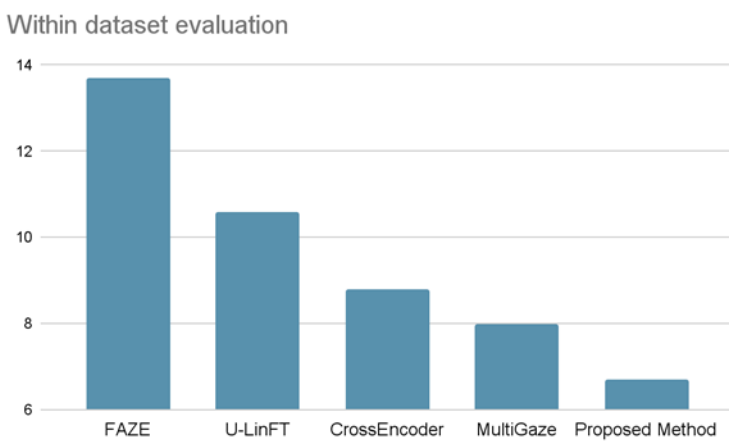

Figure 11. Evaluation comparison on GazeCapture dataset

Table 1 and Figure 11 summarize the performance comparison on the GazeCapture dataset with previous state-of-the-art gaze estimation methods. The selected comparison methods include FAZE (Park et al., 2019), U-LinFT (Yu et al., 2020), CrossEncoder (Sun et al., 2021), and MultiGaze (Gideon et al., 2022), all of which demonstrate comparable gaze estimation performance. Notably, the proposed method surpasses the performance of all comparison methods. This superiority is attributed to the proposed application of the rotation matrix in the feature space approach which enhance the model's ability to extract more robust gaze-related features. The generalization ability of this approach will be further explored in the following Chapter 4.4.2.

Cross Dataset

Table 2. Evaluation comparison on cross dataset (angular error)

Method	**Evaluation Dataset**	
	ETH-XGAZE	**EYEDIAP**
FAZE (Park et al. 2019)	25.6	24.7
U-LinFT (Yu et al. 2020)	22.8	19.6
CrossEncoder (Sun et al. 2021)	19.4	17.4
MultiGaze (Gideon et al. 2022)	16.9	16.2
Proposed Method	11.2	9.7

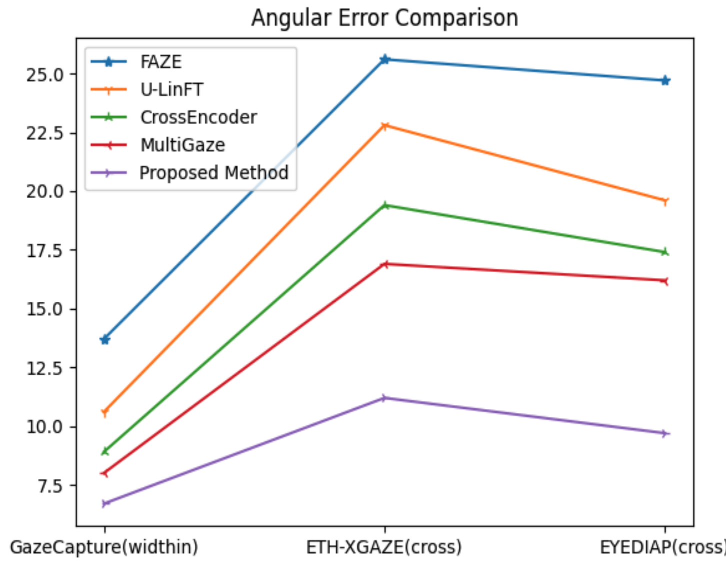

Figure 12. Evaluation comparison on cross dataset

To assess the generalization ability of the proposed method, I conducted cross-dataset evaluations. The models trained using the GazeCapture dataset are tested on different datasets, namely EYEDIAP (Mora et al., 2014) and ETH-XGAZE (Zhang et al., 2020), each characterized by slightly different domain features. Angular error measurements were employed for performance evaluation, and the results are presented in Table 2 and Figure 12.

The cross-dataset evaluation reveals notable disparities among the models. FAZE and U-LinFT exhibit subpar results which indicate limited generalizability. In contrast, CrossEncoder and MultiGaze, leveraging representation learning, demonstrate improved performance. Notably, the proposed method outperforms all previous approaches by a significant margin. This outcome clearly proves the effectiveness of the proposed approach wherein the application of the rotation matrix in the feature space. This approach contributes to increased accuracy and enhances the trained model's generalizability.

Conclusion

In this research, I proposed an unsupervised learning-based gaze estimation system for the screening of ADHD. The main objective was to introduce an innovative gaze estimation system utilizing unsupervised learning, with a distinct emphasis on addressing the challenges associated with the traditional diagnostic process for ADHD. Through a series of comprehensive experiments, I demonstrated the efficacy of the proposed gaze estimation system. The system not only outperformed previous methods, as evidenced by within-dataset evaluations on Gazecapture, but also proved its robust generalization capabilities in cross-dataset evaluations. By capitalizing on gaze abnormalities as potential biomarkers for ADHD, the proposed system presents an efficient and objective screening tool. The reliance on eye images for screening introduces a non-invasive and accessible dimension to the diagnostic process, potentially revolutionizing the way ADHD is identified and addressed.

In conclusion, this research contributes to the growing body of work aimed at leveraging advanced technologies for the betterment of healthcare diagnostics. The proposed gaze estimation system not only showcases technical advancements but also holds the potential to make a tangible impact on the early detection and intervention of ADHD. While the proposed method marks a significant step forward, there are avenues for future exploration. Further refinements in the gaze estimation model, expanded datasets encompassing diverse demographic groups, and continued validation through clinical studies can enhance the system's reliability and applicability in real-world scenarios. Additionally, future steps involve utilizing clinical data on the eye movement of individuals with ADHD to evaluate the functionality of the prototype in real-world context.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Babu, M. D., JeevithaShree, D. V., Prabhakar, G., & Biswas, P. (2019). USING EYE GAZE TRACKER TO AUTOMATICALLY ESTIMATE PILOTS'COGNITIVE LOAD. In 50th International Symposium of the Society for Flight Test Engineers (SFTE).

Cuemath. (2021, Oct 19). *"Rotation Matrix"*: Cuemath. https://www.cuemath.com/algebra/rotation-matrix/

Funes Mora, K. A., Monay, F., & Odobez, J. M. (2014, March). Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras. In Proceedings of the symposium on eye tracking research and applications (pp. 255-258).

Galgani, F., Sun, Y., Lanzi, P. L., & Leigh, J. (2009, March). Automatic analysis of eye tracking data for medical diagnosis. In 2009 IEEE Symposium on Computational Intelligence and Data Mining (pp. 195-202). IEEE.

Gideon, J., Su, S., & Stent, S. (2022). Unsupervised multi-view gaze representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5001-5009).

Goto, Y., Hatakeyama, K., Kitama, T., Sato, Y., Kanemura, H., Aoyagi, K., ... & Aihara, M. (2010). Saccade eye movements as a quantitative measure of frontostriatal network in children with ADHD. Brain and Development, 32(5), 347-355.

Kar, A., & Corcoran, P. (2017). A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. IEEE Access, 5, 16495-16519.

Kasahara, I., Stent, S., & Park, H. S. (2022, October). Look both ways: Self-supervising driver gaze estimation and road scene saliency. In European Conference on Computer Vision (pp. 126-142). Cham: Springer Nature Switzerland.

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., & Torralba, A. (2016). Eye tracking for everyone. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2176-2184).

Levantini, V., Muratori, P., Inguaggiato, E., Masi, G., Milone, A., Valente, E., ... & Billeci, L. (2020). Eyes are the window to the mind: Eye-tracking technology as a novel approach to study clinical characteristics of ADHD. Psychiatry Research, 290, 113135.

Park, S., Mello, S. D., Molchanov, P., Iqbal, U., Hilliges, O., & Kautz, J. (2019). Few-shot adaptive gaze estimation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9368-9377).

Ross, R. G., Olincy, A., Harris, J. G., Sullivan, B., & Radant, A. (2000). Smooth pursuit eye movements in schizophrenia and attentional dysfunction: adults with schizophrenia, ADHD, and a normal comparison group. Biological psychiatry, 48(3), 197-203.fv

Shah, S. M., Sun, Z., Zaman, K., Hussain, A., Shoaib, M., & Pei, L. (2022). A driver gaze estimation method based on deep learning. Sensors, 22(10), 3959.

Sun, Y., Zeng, J., Shan, S., & Chen, X. (2021). Cross-encoder for unsupervised gaze representation learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3702-3711).

Yu, Y., & Odobez, J. M. (2020). Unsupervised representation learning for gaze estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7314-7324).

Zhang, X., Park, S., Beeler, T., Bradley, D., Tang, S., & Hilliges, O. (2020). Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16 (pp. 365-381). Springer International Publishing.