Transforming Human-Computer Interaction with Real-Time Gaze-Tracking Technology

An undergraduate project proposal report submitted to the

Department of Electrical and Information Engineering
Faculty of Engineering
University of Ruhuna
Sri Lanka

in partial fulfillment of the requirements for the

Degree of the Bachelor of the Science of Engineering Honours

by

N. Hariharasakthy	-	EG/2020/3956
M.H.A. Minhaj	-	EG/2020/4076
M.N.S. Morais	-	EG/2020/4077
J. Praveenan	-	EG/2020/4119

Yugani

....
Ms. O.G.Y.N. Gamlath
(Supervisor)

Abstract

Human Computer Interaction (HCI) is the field of study which focuses on optimizing how users and computers interact. HCI can be identified as a multidisciplinary field. Gaze tracking is one of the emerging technology in HCI. Eye movement tracking is a technique use for checking the usability problems in HCI. Real – time gaze tracking enhances the user experience based on the eye-movement and its interaction.

An alternate paradigm to conventional input techniques like the mouse and keyboard is provided by eye recognition in human-computer interaction (HCI), which offers an interface that is easier to use, more accessible, more intuitive. Tasks can be performed rapidly, while ensuring dependency on traditional devices. Moreover, this approach is particularly beneficial for people with disabilities. The main reason behind is, it enables hands-free interaction with computers and digital interfaces. Additionally this is a benchmark to enhance accessibility and improve the quality of life for people with mobility impairments.

A novel approach to HCI will be proposed, by training a deep learning model. The aim is to utilize the model to identify eye-tracking in real-time and control computer applications. The primary goal of the project is to implement a software application capable of recognizing eye-movements and triggering predefined actions. The proposed system goes beyond existing applications while ensuring the reliability, accuracy, cost –effectiveness and accessibility. More specifically, this project proposes enhancement in computer controlling for assistive communication, web browsing and digital interaction, for gaming applications. Project seeks to automate a variety of significant and commonly utilized operations inside the computer environment, in contrast to current systems that frequently target single tasks. On the other hand, a user-friendly interface is suggested to trigger eye movements. The ultimate target of this project is to contribute to the advancement of HCI technologies, offering a robust and adaptable tool that uses computer vision and deep learning to improve accessibility while streamlining user contact with computers.

Acknowledgment

We would like to express our heartfelt appreciation to our supervisor, Ms.O.G.Y.N. Gamlath, Lecturer, Department of Electrical and Information Engineering, for the encouragement, guidance, and support for the progress of Our Final Year Undergraduate Project, Transforming Human-Computer Interaction with Real-time Gaze-tracking technology. We extend our sincere gratitude to our Head of the Department Dr. Rajitha Udawalpola and the module coordinator for the EE7851 Undergraduate Project, Dr. K.M.I.U. Ranaweera, Senior Lecturer, Department of Electrical and Information Engineering, for her invaluable guidance and offering immense insight and direction throughout the project. Our appreciation also goes to Department of Electrical and Information Engineering for providing us with necessary resources and facilities for the research and preparation of the project.

Contents

List of Figures

List of Tables

Acronyms

3D	Three-Dimensional
AI	Artificial Intelligence
API	Application Programming Interface
CHI	Computer Human Interaction
CNN	Convolutional Neural Network
DINN	Deep Integrated Neural Network
DTW	Dynamic Time Warping
EAR	Eye Aspect Ratio
EGT	Eye Gaze Tracking
EMDR	Eye Movement Desensitization and Reprocessing therapy
GPU	Graphics Processing Unit
HCI	Human Computer Interaction
HMI	Human Machine Interaction
IR	Infrared
MAD	Mean Absolute Deviation
MATLAB	MATrix LABoratory
ML	Machine Learning
MMI	Man Machine Interaction
MSE	Mean Squared Error
NLL	Negative Log-Likelihood
PCCR	Pupil Center Corneal Reflection
PTSD	Post-Traumatic Stress Disorder
RAM	Random Access Memory
ReLU	Rectified Linear Unit
RGB	Red, Green, and Blue
RGB-D	Red Green Blue-Depth
RMSE	Root Mean Squared Error
ROI	Region Of Interest
SSD	Solid-State Drive
UX	User Experience
VGG	Visual Geometry Group
WIMP	Windows, Icons, Menus, Pointers
YOLO	You Only Look Once

Chapter 1

Introduction

The goal of Human-Computer Interaction (Human Computer Interaction (HCI)) has always been to develop more natural and intuitive ways for people to interact with digital systems. In the domain of (HCI) Eye Tracking technology takes a unique position. Eye tracking technology can be recognized as a useful technology used in different domains such as medical diagnosis[5], marketing[6], and computer vision[7] and human-computer interaction. Traditional input devices, such as keyboards, mice have accessibility challenges individuals with physical disabilities. With respect to gaze-tracking technology has hiaghlighted as a promising approach, allowing users to manipulate computer programs with their eyes, improving accessibility and allowing hands-free engagement.

Existing eye tracking systems are sometimes expensive to purchase and inaccurate. Many users are unable to use eye tracking systems because of these restrictions. Additionally, it has made eye tracking studies difficult for scholars that are interested. As a result, numerous researchers are creating eye tracking devices that are both accessible and reasonably priced.

Deep learning and computer vision play vital roles in this topic. Recent advancements in the above sectors have notably improved the accuracy and efficiency of gaze-tracking systems. By analyzing eye movements, these systems grant users to interact with computer applications without physical contact. In addition to being more convenient, this method enables people with physical limitations to use computers on their own. But issues like precision, usability, and environmental adaptation continue to exist.

The main goal of the project is to address these challenges by developing a real-time gaze-tracking technology by addressing the current challenges. On the other hand, enhancing user-experience by providing a natural and efficient method is another key achievement through the project. The study targets to increase the potential of eye-tracking systems for contemporary computing applications by concentrating on the analysis and recognition of eye movements.

1.1 Background

1.1.1 Human Computer Interaction

The main objective of Human Computer Interaction is to enhance the way users interact with computers. Here ease of use, accessibility and responsiveness are prioritized. Traditionally. Computer interaction depends on input devices like keyboard, mouse and touch screens. However, modern HCI explored beyond these approaches, integrating alterative modalities. For example gesture recognition, speech commands, recognizing facial expressions and eye- tracking. The focus is on facilitating natural interactions with systems, which is a crucial element that is becoming more and more significant in a variety of domains related to human-computer interaction[8].

The increasing availability of computational power and artificial intelligence enable the implementation of real-time eye tracking systems. Moreover, it leads to reshape digital interaction and assistive technologies. Eye-tracking technology allows users to control applications by reviewing gaze direction .This method has profound effects in fields such as assistive communication, gaming, healthcare, and virtual reality, where gaze-based interactions enhance usability and engagement.

Ongoing advancements in computer vision and deep learning further enhance the accuracy and responsiveness of gaze-based control systems, marking them as a transforming solution for the future of human-computer interaction.

1.1.2 Gaze Tracking Technology Using Deep Learning and Computer Vision

More accurate and efficient eye movement can be analyzed through the enhanced capabilities of deep learning and computer vision. By using these technologies, computers may more accurately understand user gaze, enabling intuitive and natural human-computer interactions.

In gaze-estimation tasks,a remarkable performance is notified for deep learning models, particularly Convolutional Neural Network (CNN). These models learn complex features including pupil center detection, iris contour mapping, and gaze vector estimation from large datasets even under varying lighting conditions and different head poses. The architecture typically involves convolutional layers for feature extraction,pooling layers for spatial dimensions, and fully connected layers for precise gaze direction prediction[9]. Deep learning models are trained on large datasets. Transfer learning techniques using pre-trained models like VGG16[10] and AlexNet[11] can enhance gaze prediction accuracy, even in environments with variable lighting and background complexity.

The other major component technique is computer vision. It enables the detection and tracking essential features. Facial landmark detection can be identified as a critical thing to understand the eye position and orientation relative to a specific face. Appearance-based methods analyze the eye image to estimate gaze without definite feature extraction, reducing dependency on lighting conditions. In conclusion, these methods, combined with computer vision algorithms, facilitate robust and real-time gaze tracking.

1.1.3 Image Processing and Preprocessing

Image processing plays a significant role in the context of gaze-tracking technologies. It enhances image uniformity, the precision of eye region detection and helps to reduce background noise.To extract features like the pupil location and eye contours effective segmentation of the eye is used. This directly impacts the gaze estimation accuracy. These preprocessing steps are fundamental in computer vision tasks like gaze tracking, user interaction analysis, and assistive technologies.

A preprocessing pipeline for eye-tracking systems using appearance-based gaze estimations was implemented by Zhang et al.(2017)[9]. The process starts with illumination normalization. It reduces the effects of varying lighting conditions on the captured image. Histogram equalization techniques were used to enhance image contrast to make sure consistent image quality. Afterwards, eye region detection is performed using Haar cascades. It could accurately distinguish the eye area, reducing background noise and improving the accuracy of feature extraction. Furthermore, morphological processing eliminates minor disturbances , and the Region of Interest (ROI) is determined for the following gaze estimation tasks.

Fischer et al. (2018)[12] used a minimal preprocessing approach to balance computational efficiency with accuracy. These steps include background removal using color-based segmentation techniques in the YCbCr color space. It leads to isolate the eye area by removing unnecessary background details. Moreover, grayscale conversion helps to simplify the image,reducing computational load. It also aids in faster gaze estimation. In order to remove small artifacts without damaging the image quality, noise reduction is performed using median filters. The preprocessing pipeline ensures the robust eye detection and tracking.

By ensuring that the gaze-tracking system functions well in a range of environments and lighting circumstances, these preprocessing techniques raise the precision and dependability of eye movement analysis. The system can precisely and intuitively handle computer interfaces by accurately interpreting user gaze by reducing background noise and improving the properties of the eye region.

1.2 Problem Statement

Individuals with physical disabilities and motor impairments face significant challenges when dealing with traditional interaction methods like mouse, keyboard and touchscreens etc. Limited accessibility can be highlighted as a most talked topic for them. Therefore this leads to limited independence and restricted engagement with digital environments.While current assistive technologies, such as switch-based controls and specialized hardware, offer solutions they are often expensive, cumbersome, and lack the intuitive control which is a must for efficient computer interaction.

Additionally, current gaze-tracking systems have challenges in usability, and adaptability.For instance many existing solutions require the use of dedicated gaze tracking making gathering data on a broad scale expensive and time-consuming. Another example is some systems aim to operate in real-time at 30 frames per second, for them they necessitate high-performance hardware to function[13]. These existing system can pose noteworthy barriers to widespread adoption and usability.

And also some are expensive to purchase[14] and inaccurate[6]. Furthermore, real-time responsiveness is hampered by inconsistent eye movement detection and expensive computational requirements, which restricts the applicability of such systems for daily use.

The project targets to address these limitations by developing a real-time gaze-tracking technology that can control computer applications hands-free. By utilizing deep learning and computer vision techniques the proposed system will enhance user experience, accuracy and a cost-effective alternative to existing systems. This project explores to advance Human Computer Interaction (HCI) by creating a gaze-based user-friendly system to control various computer applications efficiently. Particularly users with disabilities to interact with computers more independently.

1.3 Objectives and Scope

1.3.1 Objectives

The main objective of this project is to develop a real-time gaze tracking system using deep learning techniques and transform Human computer Interaction through a software application. The below mentioned objectives can be highlighted.

- Find a dataset to train, design a suitable architecture and develop the model.
- Train and optimize a CNN-based model to get real-time response with high accuracy.
- Develop a gaze-tracking prototype that can detect eye movements and lead for the control for computer applications.
- Enhance gaze-based interactions for actions like selecting options, navigating interfaces, and control applications.
- By addressing the existing applications and research gaps introduce a novel approach to system control using gaze tracking technology.

- Compare the gaze-based interactions with traditional inputs.
- Ensure the project is cost-effective and compatible so it is affordable for daily use.
- Identify the technical challenges and propose solutions specially when having precision errors, calibration issues and other conditions like lighting conditions, user diversity.
- Implement a user friendly application for gaze tracking by utilizing the CNN model and computer vision and software development technologies.

1.3.2 Scope

- Development of a real-time gaze tracking system, that identifies and interprets eye movement of users to have an enhanced interaction with computer applications.
- Integrating developed deep learning model with a software application which allows user to control computer and interact with softwares using the gaze -based commands.
- Using the CNN architecture to detect different eye movements with a high precision and accuracy.
- Enhance the HCI by exploring how gaze-based controls improve usability and accessibility for both general users and individuals with physical disabilities.
- Develop the gaze tracking system with real-time response with high accuracy.
- Address the problems related to existing systems, seeking the gaps and suggest a novel contribution to Human Computer Interaction (HCI).
- Design an intuitive, hands free control system which can be used by disabled people to interact with computers effectively and efficiently.
- Consider the various conditions like lighting conditions, head positioning, age groups and including users who wear glasses to ensure robustness.
- Introduce a cost effective solution which is affordable and that can be used on ordinary laptops without depending on complex, external hardware requirements.

Chapter 2

Literature Review

2.1 Previous Work

2.1.1 Human Computer Interaction

Human Computer Interaction (HCI) is referred as Man Machine Interaction (MMI), Computer Human Interaction (CHI), and Human Machine Interaction (HMI) in various research papers.[15] The major focus of all these terms is only to provide user-friendly and effective computer-human interactions. Whenever the natural powerful processor which is human try to communicate with another powerful information processor which are computers through a limited bandwidth medium, it's called as Human Computer Interaction.[15] General Human Interactions involve natural communication methods including language, emotions, expressions, and gestures, whereas computers may vary from desktops to embedded devices. As people depend more and more on digital systems, mobile apps, and smart gadgets, HCI has emerged as a crucial component of daily life due to the quick digital evolution.[16]

WIMP-based (Windows, Icons, Menus, Pointers) interfaces were the foundation of early HCI because they allowed the interaction in a hierarchical and structured way.[15] But as technology has advanced, interaction paradigms have changed to include developing virtual environments, context-aware systems, and wearable computing, which improves usability and functionality. What a system can do, how its operations help to fulfill its goal and why it was built are referred to as its functionality.[1] In 1990s, the word usability was considered as mostly usable word which illustrates how well a system can be used to achieve particular goals of a particular user.[15] Ultimately , a user-friendly HCI design will be produced by following the well-balanced approach between usability and functionality.

Various interaction modes, such as image, audio, and sensor-based approaches, are investigated in modern HCI researches.[1] To get the better understanding of the user intent and enhance the interaction, image-based HCI utilizes the facial analysis, gesture recognition, and gaze detection. For example, eye-tracking technology enables interactions more accessible and effective by allowing users with disabilities to control systems with their gaze movements.In addition to that, Users can engage with their digital material in a natural way with the help of gesture-based interfaces, especially in virtual and augmented reality.[1][2]

By using speech recognition, speaker identification, and auditory emotion analysis, audio-based HCI improves hands-free interaction and increases the responsiveness of AI assistants such as Google Assistant, Alexa, and Siri. Sensor-based HCI opens up new possibilities along with the motion tracking, haptic feedback, and pressure-sensitive inputs. It is vastly used in virtual reality, gaming, robotics, and assistive technology. Although issues like real-time processing and input synchronization still exist, multimodal HCI improves usability and immersion by combining voice, gesture, and gaze-based interactions to get the better User Experience (UX). Future context-aware and intelligent interactions will be powered by wearable computing and adaptive AI along with the advancements in assistive technology, AI-driven interfaces, and healthcare sectors.[1]

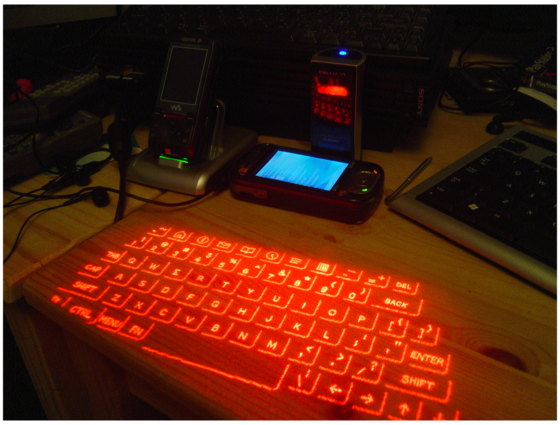

Figure 2.1: Projection Keyboard [1].

Projection Keyboard shown in the Figure 2.1 is one of the most famous laser based virtual keyboard which projects the **QWERTY-like layout** on any surface using red light. Here as the name suggests , the user can be able to give **text inputs without the need of physical keyboard** by the use of motion sensors. This HCI approach improves the user experience , accessibility and portability in mobile computing , assistive technologies and military applications.

2.1.2 Existing Gaze Tracking Technologies and Applications

In the recent researches, gaze tracking is a widely used area to understand the human behavior and advancing human interaction. Several studies have been conducted using traditional technologies to capture gaze moments with high precision, utilizing Infrared (IR) Cameras and Pupil Center Corneal Reflection (PCCR) technique. However, considering the cost of monitoring and controlled environments required for traditional methods , recent advancements have shifted towards the deep learning and computer vision field for eye movement tracking. Especially, the Convolutional Neural Networks (CNNs) has been widely utilized as the focal point in many research papers.[2][3] Other than the basic CNN model, some researches have been adopted by YOLOv3 tiny architecture of CNN with 23 layers to execute without the need of GPU accelerator.[17]

Existing Gaze Tracking Approaches

In the classical systems, notable gaze tracking research is primarily focused on the IR cameras to detect the corneal reflections and iris glints.To enhance the usability of the system, Zhu and Ji proposed a gaze tracking system in [18] which is capable of handling dynamic and natural head movements which eliminated the need for a static head position while using the eye tracking technologies. Another significant research was done by Macinnes et al. in [19] explored the balanced nature between the tracking accuracy and mobility in the context of wearable eye-tracking devices and traditional IR-based solutions. While past research in this area highlights the effectiveness of IR-based tracking, on the other hand these technologies have significant limitations due to the dependency of the specific hardware requirements and controlled conditions.

In the context of IR-based tracking, one of the most recent developments in the healthcare sector is **Wearable Eye Tracker with Mini-Infrared Point Sensors** which is evidenced in [20]. In order to monitor and arrange the treatment facilities to the post-traumatic stress disorder (PTSD) and eye movement desensitization and reprocessing therapy (EMDR) , a device has been developed by Chang et al. using two sensors which are significantly cost effective compared to other sensors to capture the both left and right eye movements. Eventually, a high correlation between the precise measurements of left and right pupils movements was achieved through this proposed approach which led to the innovations of wearable and cost effective tracking applications in the medical field.[20]

Another interesting study was conducted by Mohamed et al. to prevent the accidents causing by driver's exhaustion in [21]. In contrast to the traditional Infrared-based systems, this study introduced a method to use the smartphones (Android Samsung Galaxy S4) for the detection of the drowsiness and yawning through facial expressions and eye blinks. The system sends the alerts to the driver whenever the system detects the unusual frequent eye blinks considering the signs of the drowsiness. Here it utilizes the traditional image processing techniques through MATLAB and JAVA algorithms for the eye tracking and provides the affordable solution for the hardware limitations and controlled environment issues in the previous studies. In addition to that ,this study suggests the additional enhancements in the automated braking and parking assistance by using only the smartphone based eye tracking.[21]

Deep Learning-Based Gaze Tracking Approaches with HCI

The evolution of Artificial Intelligence (AI), Deep Learning and Computer Vision , has reduced the need of conventional IR based methods discussed above. In this context, to achieve the high accuracy under natural head movement scenarios, the **monocular 3D gaze-tracking system** which involves the geometric constraints has been proposed in [22] by Jigang et al. and it has also been extended to mobile applications. In the most recent research, considering the cost effectiveness, the **webcams were used as the input devices** to capture the eye movements to increase the accessibility by omitting the requirement of the external hardware using the Convolutional Neural Networks (CNNs).[23] The results of this research were further adopted to extend its functionality to various environmental conditions through enhancing the robustness and gaze tracking capabilities in [24].

The emerging approaches of Gaze Tracking and HCI were extensively analyzed in plenty of research papers to identify the significances of eye movements in relevance to the User Experience enhancement.A major component of usability in Human-Computer Interaction is high-quality interface design. A research [25] conducted by Poole et al., the user attention and cognitive processing were examined by using fixation duration, saccades, and gaze patterns. The results highlighted on this research were especially focused on assistive technology for those with disabilities.[25] Another research study conducted by Modi et al. in [2] emphasized the importance of gaze tracking in the marketing field by examining user behavior on social media websites using a CNN deep learning model. Through this research, they analyzed the level of eye fixations of users on social media websites, specifically using the Pepsi Facebook page by following the user engagement analysis as shown in the Figure 2.2. Based on these eye fixations, the marketing and designing teams could be able enhance the page designing ideas to attract more attention to their products.[2] This system was developed without any additional hardware setups by only utilizing the web camera that is annexed with the desktop application. Besides that there are several applications were developed utilizing the machine learning models for the consumer attention analysis in mobile advertisements by the involvement of gaze tracking.[26] This also further enhanced by Tsubouchi et al. in [27] through synchronizing the user's gaze with real time ads for personalized marketing.

As the Gaze tracking technology continues to exist in various fields, the transportation also became involved with the Gaze tracking methods for the driver safety by observing and predicting the right decision in the challenging situations of the driver by utilizing the CNN in model training in [28]. A Deep Integrated Neural Network (DINN) was created by Kanade et al.(2019) combining the Convolutional Neural Network (CNN) and a Deep Integrated Neural Network in [29]. The main objective of this study was to classifying the eye states by using the driver's photos to detect any signs of drowsiness using Deep Neural Network. Based on the amount of images we provide, it ideally detects how frequently and how long the driver blinks and it could be able to define the fatigue and drowsiness level. This introduced DINN model has trained using 2400 subjects' data and achieved the performance more than 90%.[29]

In the gamming context also gaze tracking technology research evolves in a rapid way which is evidenced by Louedec et al. in the research of [4]. As the same way of facial recognition using CNN , the individual chess pieces were recognized by utilizing the image centric behavior of CNN in this study. Here the CNN

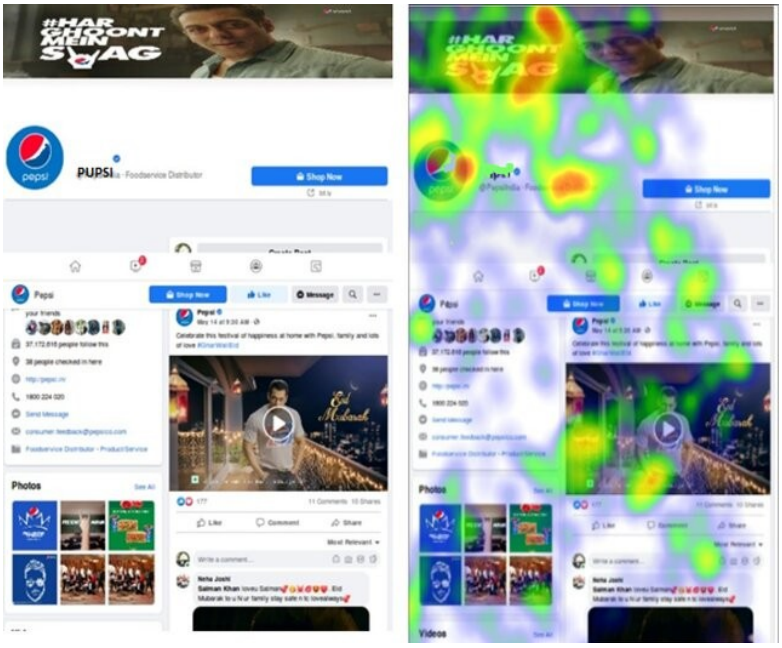

Figure 2.2: Pepsi Brand Page in Facebook and Gaze Fixation Plots [2].

model was developed by adopting the eye tracking input data of 30 chess players in different skill levels to identify the areas of chess board where a player is focusing on.[4] This study demonstrated the prediction of the player's next move with the usage of saliency map which shows the board sections.

In general, gaze tracking approaches have been expanded its applicability from conventional Gaze tracking techniques to more sophisticated deep learning and computer vision approaches. Most of the recent studies employed with the use of CNN models, wearable sensors, and smartphone-based solutions on improved precision, cost effectiveness and accessibility. In the case of wellness monitoring with medical therapy and driver safety in the transportation have huge impact by the development of gaze tracking technologies. In order to analyze the user behavior and the User Experience (UX) when interact with the user interfaces , the gaze tracking technology helps a lot in the field of the marketing, gaming and HCI. The adoption of deep learning in gaze tracking has the potentials to enhance real time interaction and decision making ability in the future researches.

2.1.3 Model Training, Validation and Testing Methods

The model training process is a important step in developing a CNN model. Before training the model, image processing should be done. The data is then prepared for the convolutional neural network (CNN) model. The architecture of research proposed by Vidhya et al (2025) is shown in Figure 2.3 after the image processing.[3] In order to extract features from input photos, this CNN model uses three convolutional layers with increasing filters (32, 64, and 128) to identify complex patterns, edges, and textures. Each convolutional layer is then processed by max pooling, which reduces spatial dimensions, retains important features, and prevents overfitting. After feature extraction, fully connected layers refine these features for prediction, with 128 and 64 neurons using ReLU activation. The output layer uses Mean Squared Error (MSE) as the loss function to predict x and y gaze coordinates without activation. The model is tuned using Adam with a learning rate of 0.001. Training uses 80% of the data, with 20% for validation, ensuring accurate gaze estimation over multiple epochs. [3]

The research conducted by MF Anzari et al (2021) on gaze tracking using unmodified camera model training process were involved two datasets: one with about 6000 images from a single person and another with images from four subjects, including variations with and without glasses. For effective weight updates, the Adam optimizer was used in the training process, with an initial learning rate of 0.00015. ReLU activation was applied to hidden layers, while Xavier initialization improved convergence by setting initial weights. The Negative Log-Likelihood (NLL) loss function, combined with the Softmax activation at the output layer, predicted probabilities for 20 gaze points. The one-eye model had three convolutional layers, one subsampling layer, and one fully connected layer; the face model had five convolutional layers, three subsampling layers, and two fully connected layers; and the both-eyes model had separate convolutional stacks for the left and right eyes, concatenated before two fully connected layers. These three network architectures were tested to handle various input types. he models were trained iteratively to minimize the loss function, refining their ability to accurately predict gaze points. These architectures were carefully selected through multiple tests to optimize performance for diverse input scenarios.[23]

The CNN model training involved collecting gaze data from 41 participants under controlled conditions in the study of [2] is done by Nandini M. et al (2022), where each user focused on calibration points displayed on a screen divided into a 4x4 grid. Each gaze captured 20 camera images, resulting in 13,940 instances after removing outliers like blinks. The CNN architecture was made up of three fully connected layers, one max-pooling layer, and three convolutional layers. To ensure an a consistent dimension, the input images were pre-processed using OpenCV. The model was trained on 70% of the dataset and evaluated on the remaining 30%. Gaze points were predicted via softmax activation at the output layer, and gradient descent was used to calculate the loss function. On the training dataset, the model's accuracy was around 84\ The CNN model training processes across the above mentioned three studies share common steps but differ in architectural choices and optimization techniques. Across these studies, the CNN models

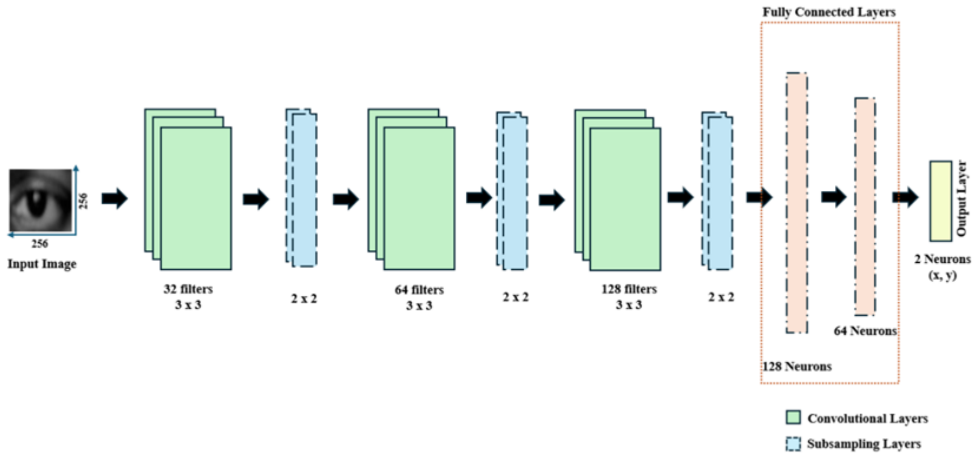

Figure 2.3: CNN Architecture for Image Processing [3].

shared key processes like convolution for feature extraction, max pooling to reduce dimensions, and fully connected layers for prediction. Variations in architecture, optimizers, loss functions, and dataset size impacted accuracy, with each approach refining techniques to enhance gaze estimation accuracy. To evaluate the CNN-based gaze-tracking model's performance on unseen data, it was trained over many epochs while its validation loss was constantly recorded. [3] [23] When the loss first decreased significantly, it was evident that the model had rapidly picked up the key characteristics required for a gaze prediction model. In the 20th epoch, the loss started to level off, indicating that the model was approaching its peak performance. [3]

The model validation and testing is the most important after model training. In the study of Nandini M et al (2022) in [2],After dividing the dataset into 70% for training and 30% for testing, the accuracy of the case analysis was 85.6%, with 84% and 83.4% achieved, respectively. Loss values were around 3.5 for training and 1.7 for testing, indicating good generalization. The model's gaze tracking accuracy was further assessed through distance error calculations using Euclidean distance and visualized with heatmaps to confirm gaze point concentration in areas of interest. Additionally, a case study with new participants validated the model's effectiveness in real-world scenarios, demonstrating reliable performance in tracking visual attention.[2]

The study done by Vidhya et al (2025) CNN model, validation was conducted using gaze tracking on images shown on a screen split into 16 grids. The CNN tracked user attention in real-time by predicting gaze locations. In order to analyze attention patterns, gaze data including fixation duration was recorded in Excel files. The Eye Aspect Ratio (EAR) was computed to measure blink rates in order to monitor engagement. Blinks were detected when EAR fell below 0.2 for consecutive frames. EAR was calculated using vertical and horizontal eye landmark distances. Blink data and timestamps were also recorded to assess user fatigue. Simulated gaze tracking was introduced by generating random gaze points inside a 200 pixel radius around the grid cell center, mimicking natural gaze variability. Heatmaps were created to visualize gaze concentration, using Gaussian blur for smoothing and contrast enhancement, with fixation probabilities calculated for each region to highlight attention distribution[3].

Second validation was done via trajectory-based evaluation. When a user follows a moving object typically a ball towards a grid on the screen, the convolutional neural network (CNN) model predicts the gaze locations of the user. All sides of the screen may be used when the user follows a preset reference route, such a circle or zigzag. The reference trajectory and the predicted gaze locations are both tracked over time, producing two sets of trajectories: one from the predetermined path and one from the model's predictions. After that, these trajectories are converted into arrays so that different performance measures may be compared. Trajectory-based accuracy was used to evaluate model performance by comparing predicted gaze paths with reference trajectory metrices like Mean Absolute Deviation (MAD), Dynamic Time Warping (DTW) and Root Mean Squared Error (RMSE) measured alignment accuracy. Additionally, a threshold-based accuracy calculation determined the percentage of gaze points falling within 150 pixels of the reference path, offering a dynamic assessment of gaze prediction accuracy.[3]

2.1.4 Image Processing Techniques

A perfect gaze tracking system depends on good image processing techniques. The quality of the pictures used to train the CNN model and the images used to extract results from the model determine its most crucial component. In image processing, different approaches play a key role in increasing the accuracy and durability of the CNN model. To enhance model generalization, picture preprocessing techniques include image resizing, augmentation, normalization, and data cleaning,.

The pre-processing pipeline for the study by Vidhya et al. (2025) begins with collecting photos of the eye area using the facial landmarks identified by Mediapipe along with OpenCV. Initially, Mediapipe is instructed to recognize 468 facial landmarks, concentrating on the area surrounding the right eye. The area of concern has been precisely captured by separating the specific places around the right eye. After being acquired, the image is resized to 256×256 pixels and converted to grayscale using OpenCV. When training a convolutional neural network (CNN), this fixed image size is essential since it ensures consistent input data, which allows for accurate predictions.[3] The dynamic image processing study was done by Nandini M. et al (2022) with the low-cost camera. For the CNN to accurately predict gaze direction, the system receives video input from a low-cost camera and processes it frame-by-frame to ensure clean, normalized, and high-quality input data. After converting each RGB frame to grayscale, histogram equalization is applied to increase contrast. Haar cascade classifiers are used to recognize the face and eyes by recognizing changes in contrast between consecutive pixels, allowing the localization of black pupils against the lighter sclera[2].

Gaussian filtering is performed to eliminate noise, giving clearer eye area pictures. The discovered eye areas are trimmed and subsequently processed to determine the pupil center using an image gradient approach and Canny Edge Detection. To enhance computational efficiency, pupil center identification is conducted on half of the frames, decreasing complexity to logarithmic time. Probability estimates are used to enhance pupil detection by integrating gradient-based computations from both eyes, enhancing accuracy. The joint probability technique improves the reliability of pupil tracking by limiting the probable pupil area using statistical measurements.[2]

In addition to these methods, various techniques have been adopted to increase the quality of input data for CNN training. For instance, the Viola–Jones classifier with Haar cascades, described by Paul V et al (2001), has been widely used to identify face features by scanning the image at several scales, assuring correct localization of the eyes [30]. This approach increases the robustness of gaze tracking by recognizing faces first and then focusing on eye areas, reducing the probability of misclassification caused by background noise or facial characteristics resembling eyes. Furthermore, Gaussian filtering is used to smooth the grayscale pictures, successfully decreasing noise and boosting the clarity of the eye areas[31]. To accurately find the pupil center, an image gradient approach is utilized, backed by Canny Edge Detection, which retains essential edge information while filtering out short-length or incorrect edges [32]. This stage refines the detection process, concentrating on the circular characteristics of the pupil, with gradients selected based on threshold values for finding the darkest parts meaning the pupil center. To improve computational efficiency, pupil center recognition is done on half of the frames, lowering complexity to logarithmic time [33]. Additionally, joint probability estimations increase detection accuracy by combining data from both eyes, limiting down the probable pupil area using statistical assessments and improving gaze estimation reliability. Finally, pixel values are normalized by scaling them between 0 and 1, ensuring constant input for the neural network and removing biases during training.

These preprocessing methods together boost the CNN's capacity to generalize by assuring clean, normalized, and high-quality input data, thereby enhancing gaze tracking accuracy and robustness in real-world circumstances.

2.2 Gaps in Literature

Gaze tracking systems for human-computer interaction (HCI) have been explored in several research papers, yet several major gaps remain in the literature that this project aims to address. Chhimpa et al.[34] propose a Revolutionizing Gaze-based Human-Computer Interaction (HCI) using Iris Tracking, that utilizes 3D eye model and geometric based analysis, which depends on mathematical models to estimate the gaze direction based on head orientation and eye position. This approach is computationally efficient but suffers from low accuracy, especially under varying lighting conditions and the presence of glasses. The involvement of Artificial intelligence, the deep learning models have been introduced to increase the accuracy and robustness of real-time gaze tracking systems.[2][3][23] The Convolutional Neural Network (CNN) is the commonly used deep learning model to extract complex patterns of eye images, making them more applicable to real-world scenarios. However, these deep learning models require large, diverse datasets for training to achieve real-time performance. Issues with dataset bias and generalization are prevalent in existing literature. Several eye gaze recognition models are trained on limited datasets that do not sufficiently represent diverse eye shapes, head positions, and lighting conditions.[22][3] This lack of diversity leads to reduced accuracy and affects the effectiveness of the model in real-world scenarios.

One major limitation is that most of the existing gaze tracking systems rely on specialized hardware devices. The Remote Eye Gaze Tracking Research was proposed by Shehu et al. in [35], that employs RGB-D and IR cameras to capture pupil and corner reflections, ensuring more accurate gaze prediction. Similarly, Event-based Human Gaze Tracking with Blink Detection is conducted by Hasegawa et al.[36], where they utilize an Event-based camera to capture pixel-level changes, providing high performance in low-light conditions with reduced power consumption. Additionally, Tobii EyeX, the popular commercial EGT device, is both expensive and requires a complex configuration setup.[37] These factors make those technologies not widely accessible for everyday uses. On the other hand, Vidhya et al.[3] and Figueroa et al.[38] have developed real-time gaze tracking systems that utilize only standard laptop webcams and deep learning model, offering low-cost alternative solution that makes gaze tracking systems with HCI more accessible and widely available. Basically, webcams capture low-resolution images, which make it difficult to extract the precise gaze direction in real-time and are extremely sensitive to changes in illumination and head movement, leading to affect the accuracy in gaze estimation.

The research taken by Khaleel et al.[39] suggests that future work should focus on extending the developed gaze tracking system to support and control a vast range of mobility applications, including mobile environments. And another significant gap in the current research is the limited scope of user-friendly software applications that use gaze detection to control multiple computer operations. Most of the studies explore technical performance (accuracy, precision, recall) without addressing software solutions.

Figure 2.4: Saliency Map Generation of Chess Game for Visual Attention [4].

Additionally, several existing solutions focus on single-use cases, such as mouse or key board controlling, visual behavior of users on Facebook page of Pepsi in real time developed by Modi et al.[2] shown in Figure 2.2, and Visual attention of chess players during game play developed by Louedec et al. [4] shown in the Figure

2.4, without providing a multitasking tool that integrates multiple functionalities into a single user-friendly platform. This research aims to address and fulfill the gaps in the literature by providing a solution that combines both software and artificial intelligence that enables the users to perform a wide range of predefined actions, such as presentation control, system navigation, and media interaction using eye movement, ensuring an intuitive and accessible user-friendly interface for individuals with disabilities.

Chapter 3

Methodology

3.1 Research Design

This section provides detailed information on different approaches utilized in this eye gaze tracking project. The resources required, method implementation, and project plan are clearly described in this chapter. In this project, we propose to develop an AI-based gaze tracking system that significantly enhances human-computer interaction (HCI), by using the webcam and deep learning model that enables the users to interact with the computers through their eye movement and is particularly useful for individuals with disabilities.

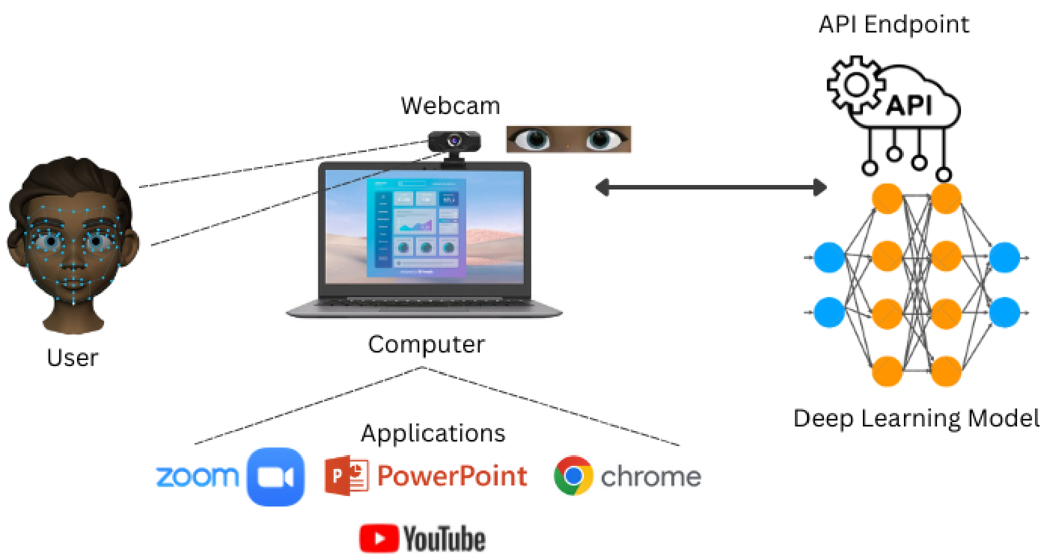

Figure 3.1: Research Design for Gaze Tracking System.

The proposed complete research design for human-computer interaction (HCI) using a real-time gaze tracking system is illustrated in Figure 3.1. In the first pre-processing input stage, the face and landmarks are captured using a webcam to extract the eye region, then normalized for consistency before being fed into the deep learning model. The extracted eye images are then sent as input to a pre-trained deep learning model. Next, in the gaze direction prediction stage, the deep learning model analyses the given input user's eye images and predicts the gaze direction and blinking actions. In the final stage, according to the predicted gaze direction or blink detection result, the system calls the appropriate API through the software application to execute the particular actions, which include controlling applications such as Zoom (mute and unmute), Media Player (play/pause, volume control), PowerPoint (slide show, navigate slides), and other system functions.

3.1.1 Design and Implementation of the Deep Learning Model

The deep learning model building is the core of the eye gaze tracking system. Dataset collection and pre-processing are an initial step of this model building. The data set of eye images will be collected from various users and different lightning conditions and head positions. Pre-processing steps include face and landmark detection using OpenCV, cropping eye region from the detected face, image normalization and resizing for model consistency, and data augmentation (brightness adjustment) to improve the model robustness.

After dataset collection and pre-processing, the selection of model architecture will be conducted based on the collected dataset properties and the intended real-time gaze tracking system. The model implementation uses the Convolutional Neural Network (CNN), which is well-suited for image-based classification problems. There are several pre-trained CNN-based model architectures available; the custom model or one of the pre-trained models will be selected based on their suitability for gaze recognition tasks, where its number of neural network layers, filters, and activation functions are customizable to achieve the specific goal of our gaze tracking project. The chosen model architecture is executed using a deep learning framework. It mainly focuses on compatibility with the collected eye dataset. Fine-tuning the model hyper parameters to improve the model's accuracy in recognizing the eye gaze.

The collected eye dataset is divided into training, test, and validation sets. It enables efficient model training and evaluation. The specified training image set is used to train the deep learning model, and model parameters are continuously updated in response to performance evaluation on the validation set, and then necessary modifications are made to improve the overall performance. Finally, the model is tested on a specified test dataset to evaluate whether it can recognize the eye gaze direction with maximum accuracy.

3.1.2 Implementation of Gaze Recognition Software for HCI

This phase involves developing the human-computer interaction based gaze tracking software application. This development process comprises two essential components: the development of the user-friendly software application and the seamless integration of a well-trained deep learning model into a real-time gaze tracking system. By combining a real-time gaze-tracking system and the trained deep learning model, the gaze-tracking software application provides a responsive interface for efficient human-computer interaction.

Implementation of User-friendly Software

The design and implementation of the user-friendly software application is an essential aspect of the real-time gaze tracking system. This software provides the interface between the user, the trained deep learning model, and other external applications, ensuring a seamless hands-free computer interaction with an intuitive experience. Processing real-time gaze data, calling the API for gaze-based actions, providing a user-friendly interface for accessibility, and guiding the users are the primary goals of this software application, ultimately enhancing the seamless hands-free computer control for individuals with disabilities.

Implementation of Real-Time Gaze Recognition System

In this phase, the previously trained deep learning model is seamlessly integrated into the real-time gaze tracking system. This integration enables the connection between the model and the webcam, allowing the real-time processing of live video streams. More specifically, algorithms are developed to capture and process each frame by frame, enabling the system to recognize face landmarks and eye regions and interpret eye movements in real time. This real-time dynamic interaction between the webcam and the trained model serves the core functionalities within the real-time gaze tracking system.

3.2 Data Collection

The data collection is the initial stage of building the gaze tracking model. The collection of datasets should be best suited and compatible for training and testing the deep learning model in the gaze tracking system. Before collecting the dataset, define a set of eye movements that the deep learning model should recognize. These include basic eye movements (left, right, center, up, down) and eye closure and opening. The main goal is to gather a balanced and diverse dataset of eye movement and blink images to ensure accurate gaze direction recognition in real-time. This caused a diligent search on the internet for pre-existing publicly available datasets. In case, an appropriate dataset might not be available on the internet, a more suitable approach is creating a custom gaze dataset. By combining the pre-existing and custom datasets, the gaze recognition model will achieve high accuracy for real-time gaze recognition in HCI applications.

Chapter 4

Timeline and Resource Required

4.1 Timeline

Table 4.1: Expected Timeline

	Feb				Mar				Apr			May					Jun				Jul			Aug				Sep				Oct				Dec
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Grouping and Selection of the project topic	■																																			
Study about the project background and Literature review		■	■	■	■																															
Submission of project proposal for supervisor's comments						■																														
Project proposal Submission							■																													
Project proposal Presentation and VIVA								■																												
Search suitable dataset and evaluate the dataset									■	■																										
Design and Implementation of Deep Learning Model architecture										■	■	■	■	■	■																					
Model Training															■	■	■	■																		
Validation and Testing of the model																		■	■	■	■															
Project Progress Report submission																					■	■														
Project Progress Presentation and VIVA																						■														
Implementation of Real-Time Gaze-Tracking Technology																							■	■	■	■										
Creating a User-friendly interface for the software																										■	■	■	■	■						
Integrate the developed software application with trained deep learning model and real-time gaze tracking system																														■	■	■	■			
Submission of draft final report for supervisor's comments																																		■		
Submission of Final report																																			■	
Final presentation and VIVA																																				■

4.2 Resource Required

Table 4.2: Resources Required and Costs

Expenses Description	**Total Price (Rs.)**
High-Performance Laptop (16GB RAM, 1TB SSD, Intel Core i7/AMD Ryzen 7)	350,000
High-End GPU (NVIDIA RTX 4070/4080 or equivalent)	400,000
PyCharm Professional	36,000 /year
Total	**786,000**

Chapter 5

Conclusion

This project highlights a novel approach to enhance Human Computer Interaction (HCI) to control various computer applications through real-time gaze tracking technology. The system integrates deep learning models like CNN and computer vision. This project addresses key limitations and gaps in existing systems, such as high cost, limited accessibility ,usability, and accuracy inconsistencies.By generalizing gaze-based interaction across applications such as computer navigation, gaming, and assistive communication, this project demonstrates its potential to significantly upgrade accessibility, particularly for individuals with physical impairments.

Additionally, the project contributes to bridging gaps in current literature by delivering a scalable, flexible and user-friendly gaze-tracking system that can be used in day-to-day tasks without depending on complex hardware requirements. Although it is still difficult to maintain consistently high accuracy, the system's focus on real-time response and usability ensures its practical viability. Future developments could focus on improving accuracy, supporting mobile platforms, and integrating other input modalities such as voice control. Overall, this project paves the way for more hands-free human-computer interaction ,fostering greater accessibility and usability for users worldwide.

References

[1] Jyoti and G. Kaur, "Research paper on human computer interaction (hci)," *International Journal for Multidisciplinary Research (IJFMR)*, vol. 5, no. 2, 2023.

[2] N. Modi and J. Singh, "Real-time camera-based eye gaze tracking using convolutional neural network: a case study on social media website," *Virtual Reality*, vol. 26, pp. 1–18, 03 2022.

[3] V. Vidhya and D. Resende Faria, "Real-time gaze estimation using webcam-based cnn models for human–computer interactions," *Computers*, vol. 14, no. 2, 2025.

[4] J. L. Louedec, T. Guntz, J. L. Crowley, and D. Vaufreydaz, "Deep learning investigation for chess player attention prediction using eye-tracking and game data," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, (New York, NY, USA), Association for Computing Machinery, 2019.

[5] P. S. Holzman, L. R. Proctor, D. L. Levy, N. J. Yasillo, H. Y. Meltzer, and S. W. Hurt, "Eye-tracking dysfunctions in schizophrenic patients and their relatives," *Archives of General Psychiatry*, vol. 31, pp. 143–151, 08 1974.

[6] M. Wedel and R. Pieters, "Eye tracking for visual marketing, now–the essence of knowledge publishing inc., boston-delft," 2008.

[7] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2176–2184, 2016.

[8] A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. S. Nayak, "Hand gesture recognition for human computer interaction," *Procedia Computer Science*, vol. 115, pp. 367–374, 2017. 7th International Conference on Advances in Computing Communications, ICACC-2017, 22-24 August 2017, Cochin, India.

[9] X. Zhang, Y. Sugano, and A. Bulling, "Evaluation of appearance-based methods and implications for gaze-based applications," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, (New York, NY, USA), p. 1–13, Association for Computing Machinery, 2019.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[12] T. Fischer, H. Chang, and Y. Demiris, "Rt-gene: Real-time eye gaze estimation in natural environments," in *Computer Vision – ECCV 2018* (V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, eds.), Lecture Notes in Computer Science, pp. 339–357, Springer, Oct. 2018. The European Conference on Computer Vision (ECCV), 2018 ; Conference date: 08-09-2018 Through 14-09-2018.

[13] E. Iacobelli, V. Ponzi, S. Russo, and C. Napoli, "Eye-tracking system with low-end hardware: Development and evaluation," *Inf.*, vol. 14, p. 644, 2023.

[14] D. Cazzato, M. Leo, C. Distante, and H. Voos, "When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking," *Sensors*, vol. 20, no. 13, 2020.

[15] R. Kumar, V. Jaiswal, and V. Nishad, "Human-computer interaction (hci)," *International Journal of Engineering Research & Technology (IJERT)*, vol. 09, no. 05, 2021.

[16] H. A. Kheder, "Human-computer interaction: Enhancing user experience in interactive systems," *Kufa Journal of Engineering*, vol. 14, pp. 23–41, October 2023.

[17] W. Kurdthongmee, P. Kurdthongmee, K. Suwannarat, and J. K. Kiplagat, "A yolo detector providing fast and accurate pupil center estimation using regions surrounding a pupil," *Emerging Science Journal*, vol. 6, p. 985, October 2022.

[18] Z. Zhu and Q. Ji, "Eye gaze tracking under natural head movements," vol. 1, pp. 918– 923 vol. 1, 07 2005.

[19] J. J. Macinnes, S. Iqbal, J. Pearson, and E. N. Johnson, "Wearable eye-tracking for research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices," *bioRxiv*, 2018.

[20] S. Chang, J.-Y. Lee, S.-I. Kim, S.-Y. Lee, and D. Lee, "Simple wearable eye tracker with mini-infrared point sensors," *IEEE Access*, vol. PP, pp. 1–1, 01 2024.

[21] M. A. Mohamed, A. I. Abdel-Fatah, and B. M. El-Den, "Fighting accident using eye detection for smartphones," *International Journal of Engineering Research and Applications (IJERA)*, vol. 4, pp. 67–73, August 2014.

[22] L. Jigang, B. Francis, and D. Rajan, "Free-head appearance-based eye gaze estimation on mobile devices," pp. 232–237, 02 2019.

[23] M. F. Ansari, P. Kasprowski, and M. Obetkal, "Gaze tracking using an unmodified web camera and convolutional neural network," *Applied Sciences*, vol. 11, no. 19, 2021.

[24] J. Singh and N. Modi, "A robust, real-time camera-based eye gaze tracking system to analyze users' visual attention using deep learning," *Interactive Learning Environments*, vol. 32, pp. 1–22, 06 2022.

[25] A. Poole and L. Ball, *Eye tracking in human-computer interaction and usability research: Current status and future prospects*, pp. 211–219. 01 2006.

[26] W. Xie, M. H. Lee, M. Chen, and Z. Han, "Understanding consumers' visual attention in mobile advertisements: An ambulatory eye-tracking study with machine learning techniques," *Journal of Advertising*, vol. 53, pp. 1–19, 11 2023.

[27] K. Tsubouchi, K. Taoka, K. Ikematsu, S. Yamanaka, K. Narumi, and Y. Kawahara, "Eye-tracking ad: Cutting-edge web advertising on smartphone aligned with user's gaze," pp. 469–474, 03 2024.

[28] P. Kanade, F. David, and S. Kanade, "Convolutional neural networks (cnn) based eye-gaze tracking system using machine learning algorithm," *European Journal of Electrical Engineering and Computer Science*, vol. 5, no. 2, pp. 36–40, 2021.

[29] L. Zhao, Z. Wang, G. Zhang, Y. Qi, and X. Wang, "Eye state recognition based on deep integrated neural network and transfer learning," *Multimedia Tools and Applications*, vol. 77, pp. 19415–19438, 2018.

[30] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," vol. 1, pp. I–511, 02 2001.

[31] M. M. Rahman, M. S. Islam, M. K. A. Jannat, M. H. Rahman, M. Arifuzzaman, R. Sassi, and M. Aktaruzzaman, "Eyenet: An improved eye states classification system using convolutional neural network," *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, pp. 84–90, 2020.

[32] R. Valenti and T. Gevers, "Accurate eye center location through invariant isocentric patterns," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, pp. 1785–98, 09 2012.

[33] K. Ahuja, R. Banerjee, S. Nagar, K. Dey, and F. Barbhuiya, "Eye center localization and detection using radial mapping," pp. 3121–3125, 09 2016.

[34] G. R. Chhimpa, A. Kumar, S. Garhwal, Dhiraj, F. Khan, and Y.-K. Moon, "Revolutionizing gaze-based human–computer interaction using iris tracking: A webcam-based low-cost approach with calibration, regression and real-time re-calibration," *IEEE Access*, vol. 12, pp. 168256–168269, 2024.

[35] I. Shehu, Y. Wang, A. Mohamed, and X. Fu, "Remote eye gaze tracking research: A comparative evaluation on past and recent progress," *Electronics*, vol. 10, p. 3165, 12 2021.

[36] S. Hasegawa, N. Sakata, and V. Sundstedt, "Learning-based event-based human gaze tracking with blink detection," 2024.

[37] A. Stolińska, M. Noga, T. Grabiński, A. Surowiec, and W. Byrski, "Usability and performance of selected configurations of eye tracking sets in augmentative and alternative communication–case study," *Zeszyty Naukowe Wyższej Szkoły Ekonomii i Informatyki w Krakowie*, no. 18, pp. 123–144, 2022.

[38] S. Figueroa, I. Pineda, P. Vizcaino, I. Reyes-Chacón, and M. E. Morocho-Cayamcela, "A webcam artificial intelligence-based gaze-tracking algorithm," pp. 228–235, 01 2024.

[39] A. H. Khaleel, T. H. Abbas, and A.-W. S. Ibrahim, "Best low-cost methods for real-time detection of the eye and gaze tracking," *i-com*, vol. 23, no. 1, pp. 79–94, 2024.