Development of a Measurement System based on Level of Interest for Providing Human-friendly Services

Hu Jiaren, Jinseok Woo† [0000-0002-6118-6617], and Yasuhiro Ohyama

Department of Mechanical Engineering, School of Engineering,
Tokyo University of Technology,
1404-1 Katakuracho, Hachioji Tokyo 192-0982, Japan,
†Corresponding author
e511904763@edu.teu.ac.jp, woojsk@stf.teu.ac.jp, ohyama@stf.teu.ac.jp

Abstract. Research on neural networks has made significant progress since the invention of the perceptron in 1957. In recent years, advancements in hardware technology have enabled the development of multilayer neural networks, commonly referred to as deep learning. These large-scale learning models have found applications across various fields, including generative dialogue artificial intelligence, image processing, and object recognition, profoundly transforming these domains. The use of these revolutionary technologies is expected to lead to the development of many robotic systems that can assist and benefit people in various ways. This study aims to advance this field further by exploring the possibility for artificial intelligence to understand human conditions and behavior, and suggest or carry out the actions that a specific one might wish to perform. To this end, we have developed a gaze measurement system in the form of glasses. The primary objective of this research is to track a specific user's gaze, identify the object of their focus, and estimate their desired actions based on their level of attention and interest in the object.

Keywords: User's gaze · Level of interest · Human robot interaction.

1 Introduction

Neural networks are models that mimic the neural circuits of the brain, capable of learning patterns from numerous inputs and predicting outputs [2]. The perceptron was proposed in 1957 [7], and subsequently, multilayer neural networks using backpropagation emerged in the 1980s [6]. Since entering the 21st century, the dramatic progress in deep learning has been facilitated by the increased computational power of computers and the availability of large-scale datasets [8].

The purpose of this research is it might be possible for artificial intelligence to understand the human condition or behavior, and the situation around this specific person, and suggest or carry on the actions that person might want to perform by learning this human's behavior or any other way can learn to analysis human's behavior. As we know, the information from the eyes is about 8 in 10 in all the ways we get information from our senses, and eye contact is also the main way we communicate without language. Therefore, among the analysis targets for user behavior, we paid attention to the relationship between gaze and level of interest. In this research, we focus on how the user's gaze moves in action patterns, such as turning lights on and off. We study how people use their gaze to recognize actions like turning lights on or off [1].

The structure of this paper is as follows. In section 2, we explain the whole system for providing human-friendly service's building and overview. Section 3 explains the hardware device and the software system for analyzing the user's gaze. Section 4 explains the main system and method in gaze analysis applied to the proposed device. Section 5 discusses experimental results to evaluate the proposed experiment. Section 6 presents the conclusions and future works.

2 A System for Providing Human-Friendly Services

There are about 80% of the information we receive comes through our eyes, making them our primary source of information [4]. The movements and focus of our eyes often show an interest in or intention of something. For example, when we want to turn off a light, many of us look towards the switch, either to leave or enter a room, and we often look at the light after the switch to confirm whether it's been turned off. This action pattern can be an indicator of our intention to control the light, which a robot could learn as shown in Fig. 1. When the same pattern occurs again, the robot can estimate and perform the desired action. This method can be applied across a wide range of applications.

For example, in restaurants, service robots are already widely deployed in Japan, and this technology could be further enhanced by analyzing and responding to each customer's behaviors and preferences [10]. It could lead customers to their tables upon arrival, offer water or desserts after meals based on observed preferences, and more. Similarly, in hospitals or nursing facilities, which face challenges in human resources, this system could enable smarter and more personalized services [9]. By understanding each person's specific behaviors, care robots or smart devices within facilities could provide more accurate and responsive care.

Creating human-friendly services requires a deep understanding of human needs and preferences. This research focuses on analyzing the human gaze, which reveals levels of interest and attention. By studying where and for how long individuals look at specific objects or areas, we can infer their intentions and interests. This analysis informs the design of systems that estimate user needs and suggest or carry out the actions.

2.1 Eye Tracking System for Analyzing Interest Levels

In order to analyze interest levels, it is crucial to accurately determine the focal points of the user's gaze. Traditional eye-tracking systems like MediaPipe typically require a clear view of the face to locate the eyes for tracking purposes [5].

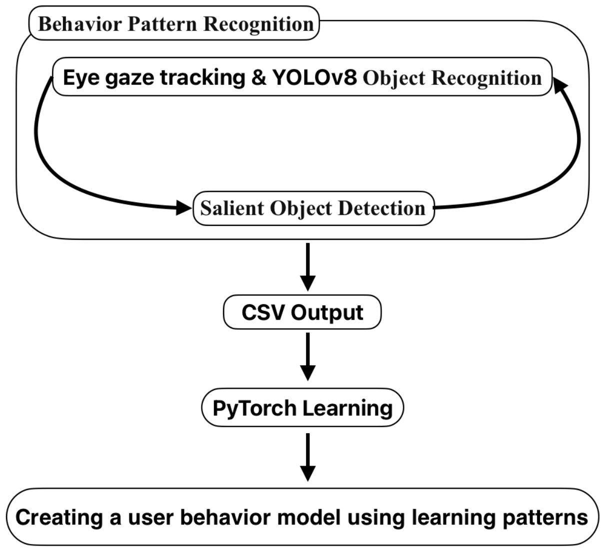


Fig. 1. action Pattern of Measurement Model

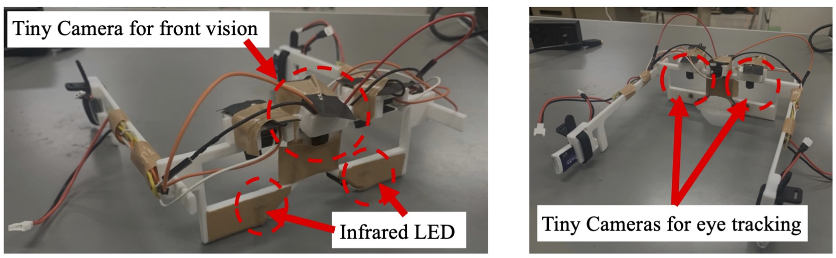
However, in scenarios where the user is wearing eyeglasses that obstruct this view, such as in our research, alternative methods must be devised. Our approach involves a custom eye-tracking system designed specifically for scenarios where conventional methods are impractical. The system leverages computer vision techniques to detect and track the pupil's position within the eyeglasses. Here, we present a method based on image processing and edge detection to achieve this goal. Therefore, in this research, we proposed an eye gaze measurement system as shown in Fig. 2. And the parts used this system are as shown in Fig 3.

The core of our proposed system involves:

- Pupil Detection Algorithm: We employ image processing techniques to isolate the pupil within the eyeglasses' lens. This includes:
  - Binary Image Conversion: Conversion of the image into a binary format to highlight areas, specifically targeting the pupil's reflection from an infrared LED appearance.
  - Edge Detection: Using the Canny edge detection algorithm helps clarify the edges in the binary image, which accurately outlines the boundaries of the pupil's reflection from an infrared LED.
  - Hough Circle Transform: Employing the Hough Circle Transform to detect circular shapes within the processed image, thereby identifying the location and radius of the pupil's reflection from an infrared LED.
- Main Objective: The main goal is to locate the reflection from the LED to identify the pupil accurately.

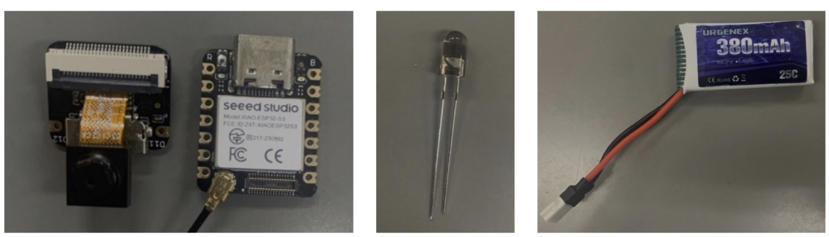
– Real-Time Implementation: The system is implemented to operate in real-time, capturing and processing frames from a camera feed where the user's eyeglasses are worn.

This methodology allows us to track the gaze direction of the user effectively, even when traditional methods are not feasible due to occlusions caused by eyeglasses. By understanding where the user directs their gaze, our system can analyze their interest levels in real-time applications.



(a) Front side of made glasses (b) Back side of made glasses

Fig. 2. The made glasses for gaze recognition



(a) Microcontroller: XIAO ESP32S3 (Sense) (b) Infrared LED (c) Lithium polymer battery

Fig. 3. Components used in gaze recognition systems

2.2 Analysis of Level of Interest

To analyze the level of interest, analyzing gaze patterns is essential. Gaze analysis helps identify what captures a user's attention and for how long, providing insights into their preferences and priorities. This data can be used to tailor responses and actions, making interactions more engaging and relevant. To analyze the level of interest, analyzing gaze patterns is essential. Gaze analysis helps identify what captures a user's attention and for how long, providing insights into their preferences and priorities. This data can be used to tailor responses and actions, making interactions more engaging and relevant.

The system for analyzing gaze patterns involves several key components:

- **Data Collection and Processing:** Collects gaze data in real-time, recording where and when users focus their attention and matching with the front camera. In Fig. 5 the user is staring at the specific position of the picture from the front camera. In Fig. 6 (a)−(c) is the situation user is staring at the center point of the picture. In Fig. 6 (d)−(f) is a scene looking at the bottom left of the figure.Each of the pictures of Fig. 6 (g)−(I), (j)−(l), and (m)−(o) shows a scene where calibration is being performed while looking at the top-right, top-left, and bottom-right of the picture, respectively. And finally, we can know the area and the proportion of the gaze when the camera can recognize the object that the user is staring at Fig. 6.
- **Gaze Pattern Analysis:** Applies algorithms to create and interpret gaze patterns, identifying areas of interest and attention durations. In this research, we set the action pattern by looking and recognizing the state of light and switching an average of 1.5 seconds 82 times, in Fig. 7 (a) the human recognizes the light on, and at the same time (b) the camera also recognizes the light is on, in (c) human heading to turn the switch off, at the same time in (d) camera can recognize the switch and the user is looking at it. When the user wants to turn the light off, in (e) the user recognizes the light is off, and at the same time, (f) the camera also recognizes the light is off, and (g) if the user wants to turn the switch on (h) the camera can know the user is looking at the switch as shown in Fig. 7. Finally, the data will be saved as CSV files.
- **Eye Tracking Device:** Utilizes advanced sensors to capture precise eye movements and fixations in Fig 6.
- **User Intent Inference:** Infers user intentions based on gaze behavior, predicting preferences and priorities.
- **Response Tailoring:** Adapts system responses and actions based on inferred user interests, enhancing user engagement and satisfaction.

This systematic approach enables the creation of user-centric services that anticipate and respond to user needs effectively. By understanding and responding to human gaze patterns, systems can enhance user experience and satisfaction across various applications.

3 The Specifications of Eye Tracking System

The eye-tracking system consists of several key components, including hardware for capturing gaze data and software for processing and analyzing this data.

- Microcontroller, Camera Module set: XIAO ESP32S3 Camera Module
- Led: infra-red rays led
- Mounting Frame: made glasses frame

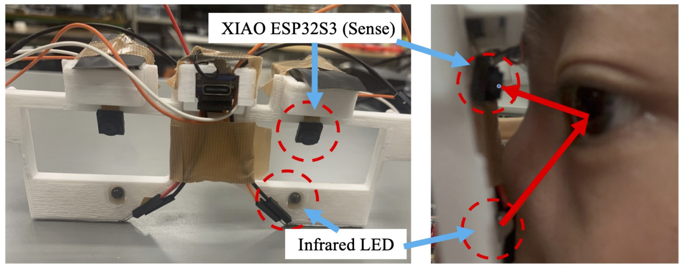


Fig. 4. The method of gaze measurement

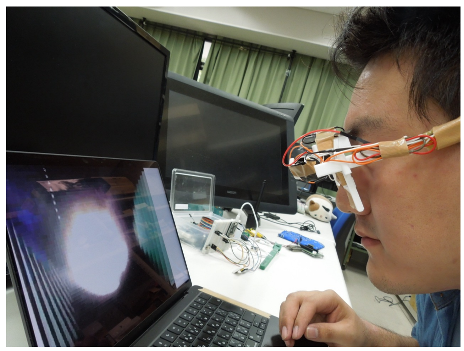


Fig. 5. Initializing gaze area

- Computer: For data processing and analysis
- Battery and Cables: To power the system and connect components

In this study, XIAO ESP32S3 (Sense) is used as a system for tracking the user's eyes. The specifications of this system are shown in Table 1.

The specifications of the computer used for learning are shown in Table 2. Since the resolution of the camera is limited, in this study, the line of sight is measured by analyzing the movement of reflected light using infrared LEDs. The specifications of the LED used in this system are as shown in Figs 4 and 5 (Table 3).

This entire system can be used wirelessly with the battery specified in Table 4. The system's operational time is approximately from 1 to 2 hours.

4 Proposed Method for Gaze Analysis

To analyze gaze data and predict user interests, we propose a method that integrates advanced object recognition techniques with neural networks. The approximate sequence of the experiment is as follows.

- **Object Recognition with YOLO [3]**

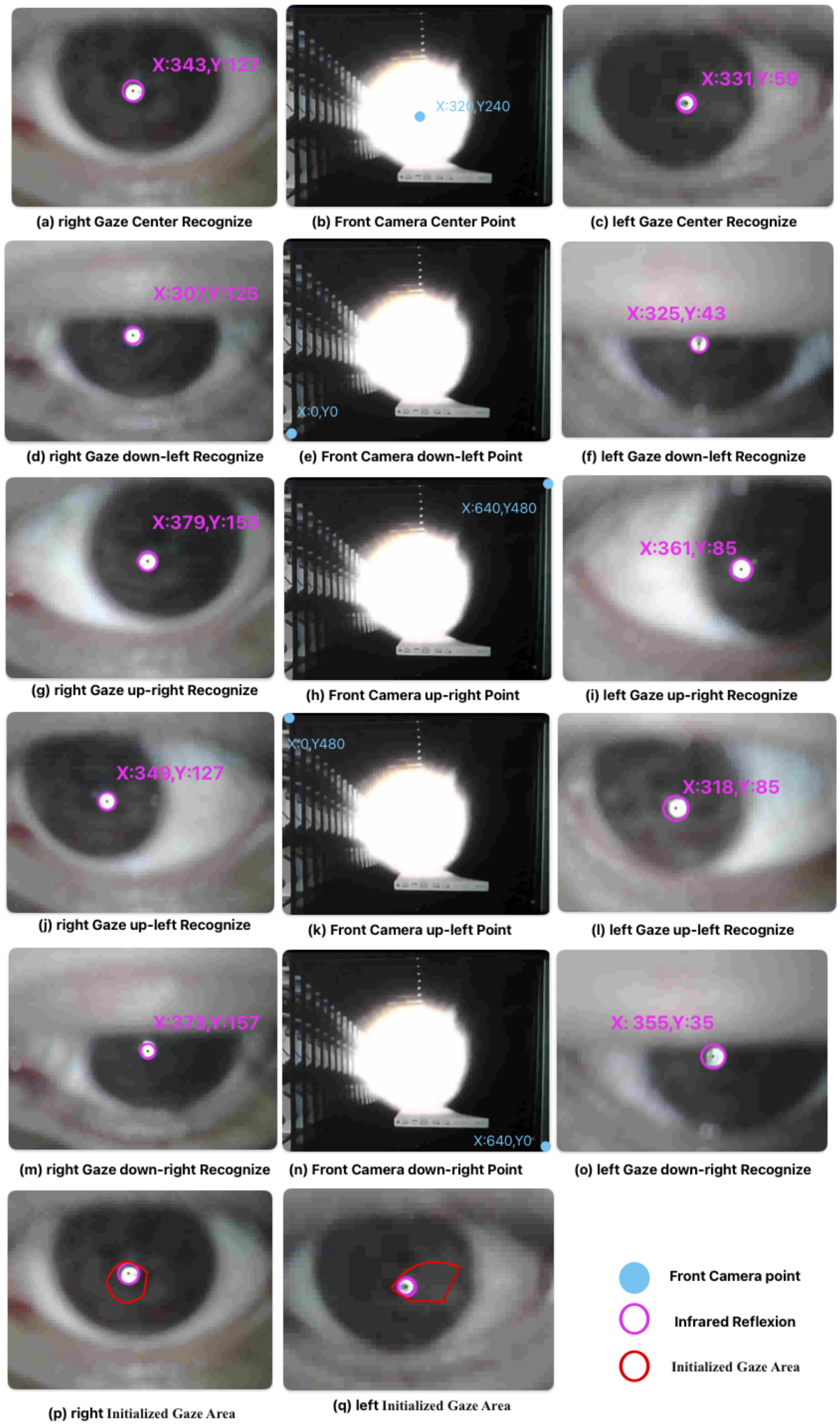
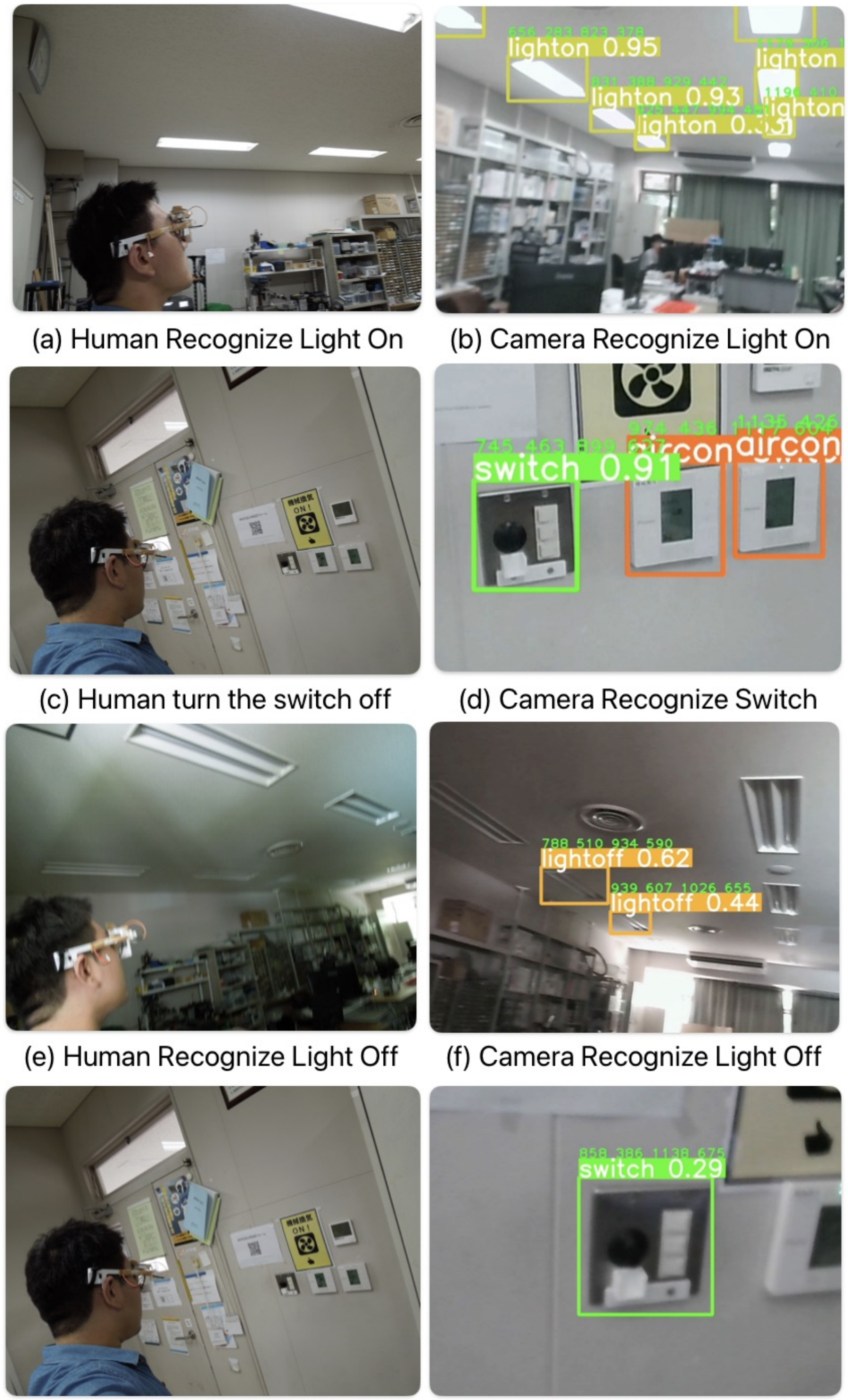


Fig. 6. Method of measuring direction of gaze



(a) Human Recognize Light On (b) Camera Recognize Light On

(c) Human turn the switch off (d) Camera Recognize Switch

(e) Human Recognize Light Off (f) Camera Recognize Light Off

(g) Human turn the switch On (h) Camera Recognize Switch

Fig. 7. Configuring the behavior pattern to be implemented in the system

Table 1. Specification of XIAO ESP32S3 (Sense)

CPU	ESP32-S3 Xtensa 32bit LX7 Dual core 240MHz
GPIO	11
ADC	9 (12bit, 0 – 4095, 3.3V)
PWM	11 (8bit 0 – 255)
UART	1 COM Serial and GPIO43(TX), GPIO44(RX)
I2C	1
SPI	1
WiFi	2.4GHz
Bluetooth	5.0

Table 2. Specification of Macbook pro M13

CPU	Apple M3 pro
Memory	18 GB
macos	15.0

  - **Data Collection:** Images of light switches and fixtures are collected from various angles.
  - **Data Preprocessing:** Images are normalized and formatted for YOLO.
  - **Model Training:** YOLO is trained to recognize the specific objects in the dataset.
  - **Evaluation and Adjustment:** The model is tested, and adjustments are made to improve accuracy with 146 pictures and 438 recognized data in YOLO. Some examples are shown in Fig. 8.
- **Data Collection and Processing**
  - **Gaze Data Collection:** The system captures real-time video from the camera mounted on the glasses frame.
  - **Image Processing:** The captured video frames are processed to identify the pupil and track its movement. This involves(Fig. 6)):
    * Binarizing the video frames to isolate potential pupil candidates.
    * Using edge detection to highlight the pupil's contour.
    * Applying Hough transformation to detect the circular shape of the pupil.
    * Extracting the center coordinates and radius of the detected pupil.
  - **Gaze Measurement:** The pupil center coordinates are used to estimate the gaze direction.
  - **User Interaction:** The gaze direction is displayed in real-time for user feedback.
  - **Light Source for Pupil Detection:** An infrared LED is used to create a circular reflection in the eye, making pupil detection easier and more accurate.
- **Integration of Gaze and Object Data**

Table 3. Specification of Infrared LED

Part NO.	OSI5LA5113A
Voltage	1.35-1.6V
Lllumination angle	15deg
Peak wavelength	940nm

Table 4. Specification of battery

Battery Type	Rechargeable Lipo Battery Pack
Voltage	3.7V
Capacity	380mAh
Discharge Rate	25C
Weight	10.5g

- **Gaze Data and Object Recognition Integration:** The system combines gaze data with object recognition results to identify which objects the user is focusing on.
- **Interest Prediction:** By analyzing the duration and frequency of focus on specific objects, the system predicts the user's interest and potential actions they might want to perform.

Through the above experiment, the level of interest in the object is confirmed, the object that the user is paying attention to is analyzed, and the user's overall level of interest is ultimately determined.

5 Experimental Results

In this study, we constructed a model using gaze tracking and deep learning to understand human states and estimate their intentions. Experimental results confirm the effectiveness of the proposed model. As shown in Fig. 9, it is possible to determine the status of key objects using real-time images of the environment, and the gaze direction information obtained from the glasses can be analyzed as shown in Fig. 6. By analyzing the object information obtained here and the gaze information together, it can be seen that it is possible to determine the object the user is looking at.

In this study, we used the PyTorch framework to train a model. The model includes a single fully connected layer with an input feature dimension of 1 and an output representing a single probability value. During training, we employed a binary cross-entropy loss function and the Adam optimizer with a learning rate set to 0.001. After 10 epochs of training, the model's accuracy on the test set was calculated and reported in Fig. 10. Fig. 10 represents the loss state and should show a value less than 1.0. When analyzing this result, the lowest loss was about 1.35 due to insufficient data volume. Fig. 10 represents the accuracy, which was ultimately improved to 50%. However, additional experiments are needed in the future due to the lack of data.

(a) Switch training data (b) Light Off training data (c) Light On training data

Fig. 8. YOLO Object Recognize training data

(a) Switch recognition results for lighting control (a) Lighting status recognition result

Fig. 9. Switch recognition results for lighting control

6 Conclusions

In this study, we have explored the potential of artificial intelligence to understand human conditions and behaviors and to suggest or perform actions based on this understanding. Our research focuses on developing a gaze-tracking system embedded in glasses to achieve this goal. By studying how people look at things and how interested they seem, we've shown that we can figure out what they're likely thinking about. Our system tracks where and for how long users look, helping us understand their interests and intentions in real-time situations. The eye-tracking system we designed uses advanced image processing techniques, like detecting pupil reflections using infrared LEDs and edge detection algorithms. This allows us to track gaze accurately, even when users wear glasses. Future research aims to improve our model's accuracy and reliability with more datasets, and real-time gaze tracking, and hope to use it practically in various fields. By using these AI and gaze tracking, we aim to create human-friendly services that adapt to each preference and behavior. This research sets the stage for developing human-friendly service systems that enhance user experience in different applications.

Acknowledgements This work was supported by JSPS KAKENHI (Grant-in-Aid for Early-Career Sci- entists) Grant Number 23K17261.

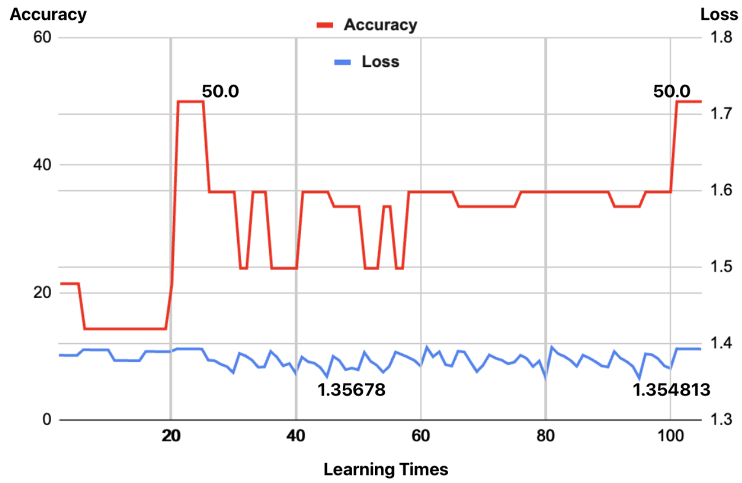


Fig. 10. The learning result of estimating the state in which the user will turn on the light by pushing the switch

References

1. Ajanki, A., Hardoon, D.R., Kaski, S., Puolamäki, K., Shawe-Taylor, J.: Can eyes reveal interest? implicit queries from gaze patterns. User Modeling and User-Adapted Interaction **19**, 307–339 (2009)
2. Arbib, M.A.: The handbook of brain theory and neural networks. MIT press (2003)
3. Hermens, F.: Automatic object detection for behavioural research using yolov8. Behavior Research Methods pp. 1–24 (2024)
4. Hutmacher, F.: Why is there so much more research on vision than on any other sensory modality? Frontiers in psychology **10**, 481030 (2019)
5. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.L., Yong, M., Lee, J., et al.: Mediapipe: A framework for perceiving and processing reality. In: Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR). vol. 2019 (2019)
6. Rojas, R., Rojas, R.: The backpropagation algorithm. Neural networks: a systematic introduction pp. 149–182 (1996)
7. Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization (1958) (2021)
8. Shinde, P.P., Shah, S.: A review of machine learning and deep learning applications. In: 2018 Fourth international conference on computing communication control and automation (ICCUBEA). pp. 1–6. IEEE (2018)
9. Wada, K., Shibata, T., Saito, T., Tanie, K.: Effects of robot-assisted activity for elderly people and nurses at a day service center. Proceedings of the IEEE **92**(11), 1780–1788 (2004)
10. Yu, Q., Yuan, C., Fu, Z., Zhao, Y.: An autonomous restaurant service robot with high positioning accuracy. Industrial Robot: An International Journal **39**(3), 271–281 (2012)