Dissertation Title:

EmoBox: Enhancing Emotional Communication for the Visually Impaired through Machine Learning

Jiaxin Liang

22018177

MSc Creative Computing

University of the Arts London: Creative Computing Institute

Supervisor: Jasper Zheng

Submission Date: Novenber 24, 2023

Video: https://youtu.be/f37EYiDYEj4

Abstract

This project leverages machine learning to enhance emotional communication for visually impaired individuals. I optimized the dataset and integrated a real-time emotion recognition system, with experimental results demonstrating its effectiveness in practical scenarios. However, challenges persist in handling subtle facial expressions and specific posture restrictions. Through this project, my goal is to address the social communication challenges faced by visually impaired individuals using advanced machine learning techniques, providing them with greater convenience and opportunities.

Acknowledgement

Over the past year, I've been incredibly fortunate to cross paths with remarkable individuals in my life! A heartfelt thank you to all the teachers who have imparted their knowledge and guidance, shaping the person I am today. My appreciation extends to my fellow classmates, and I am deeply grateful for the unwavering support of my parents, who have consistently encouraged me to pursue my passions!

Content

Chapter 1

Introduction

In the contemporary landscape, the realms of artificial intelligence (AI) and machine learning (ML) have notched remarkable strides, ushering in myriad conveniences to our daily lives and professional domains[3]. The merits of AI and ML unfold across several dimensions: foremost, their automation and heightened efficiency markedly amplify work productivity, alleviating the burdens borne by individuals; secondly, the precision and reliability inherent in AI and ML models, trained on extensive datasets, endow them with a heightened accuracy and stability within specific domains; moreover, their intelligent decision-making capabilities empower these technologies to furnish us with judicious decision recommendations by scrutinizing vast datasets; lastly, the capacities for continuous learning and adaptability enable ML algorithms to dynamically evolve and adjust based on new data, showcasing commendable flexibility. Consequently, I will unveil my culminating project, an initiative aimed at transmuting facial emotion recognition into Braille, thereby endowing the visually impaired with the ability to "sense" the expressions of those they engage with through the tactile medium of Braille. This inventive mode of communication harmoniously amalgamates the strengths intrinsic to the fields of AI and ML.

The procedural realization of the project unfolds in three sequential steps: firstly, the utilization of facial recognition technology to capture the facial features of individuals engaged in communication with the visually impaired; subsequent to this, the application of ML algorithms to transmute these features into specific emotion categories such as joy, sadness, anger, etc.; and ultimately, the translation of this emotional information into Braille, facilitating the visually impaired in perceiving the emotional states of their communication counterparts through the sense of touch. In summation, this project serves as a testament to the immense potential of AI and ML in ameliorating the quality of life for the visually impaired, providing them with an enriched social milieu. This exemplification vividly underscores the pragmatic application value of AI and ML technologies in resolving real-world predicaments.

1.1Motivation

In an eccentrically conceived restaurant in the United Kingdom (Dans le Noir?), I found myself immersed in an hour-long sojourn through a realm of absolute darkness (refer to Figure 1). On that particular day, my compatriot and I ventured into this enigmatic domain, seated vis-à-vis, with our mutual existence discernible solely through the language of touch. Amidst this fleeting gustation, I grappled with feelings of trepidation, anxiety, and an earnest yearning for illumination. Foremost among these concerns was

the incapacity to discern the expressions of my conversational counterpart seated across from me. Despite the audibility of her voice and the palpability of her touch, the inability to visually perceive her expressions during our discourse left me disconcerted, impeding my ability to gauge her emotional nuances.

figure 1

This encounter prompted profound contemplation: Could there exist a mystical solution enabling us to "see" the expressions of conversational counterparts even in the absence of sight? In pursuit of an answer, I embarked on a preliminary inquiry, discovering that visually impaired individuals within my Chinese familial circle often grapple with despondency due to their visual limitations. While the market hosts a plethora of ML-driven products and functionalities catering to the convenience of individuals with disabilities—ranging from color filters designed for specific types of color blindness to web pages that transmute text or images into speech and applications adept at adjusting images to facilitate contour recognition—it appears that a product enabling visually impaired individuals to "see" the expressions of their conversational counterparts is yet to materialize.

This challenge impelled me to delve into the boundless prospects within this domain, seeking innovative technologies capable of illuminating the lives of the visually impaired.In my conceptualization, this tool ought to be unintrusive to conversation, devoid of auditory output, and rely on tactile cues to convey the expressions of conversational counterparts.

1.2 Objectives

To encapsulate, I have delineated the objectives of this project: to construct, grounded in machine learning, a tool that empowers individuals to "see" the expressions of conversational counterparts in scenarios bereft of visual acuity. The tool, employing a camera, captures facial expressions and translates recognized expressions into touchable Braille output. The objectives encompass: 1. User-friendly operation, 2.

Real-time capture of facial expressions, 3. Sustained precision in expression recognition for Asian faces, 4. Design centered around user experience, and 5. Interactive engagement with actual users.

Chapter 2

Background

2.1 Those Plunged into Darkness

According to data from the World Health Organization, there are currently over 2.2 billion people worldwide with impaired vision or blindness, with over 1 billion of them facing untreated issues such as myopia, hyperopia, glaucoma, and cataracts[1]. Statistics released by the World Blind Union in 2022 reveal that, globally, millions of books are published each year, yet only 1% to 7% of these books are accessible to the 285 million blind and visually impaired individuals, with 90% of them belonging to the low-income populations of developing countries[4]. Notably, China boasts the highest number of blind individuals globally, with over 17 million people, implying that approximately one in every eighty individuals in China is visually impaired. However, the availability of products specifically designed for the blind and visually impaired is remarkably limited.

Nevertheless, the blind are social beings with psychological needs for recognition and acceptance. To create truly useful products, it is essential to delve into both the surface and latent needs of the blind, designing solutions that meet these needs. Through analysis, three psychological characteristics of the blind emerge. Firstly, there is a duality of loneliness and social interaction. Those with non-congenital visual impairment may fear crowds, reducing their social interactions and leading to feelings of loneliness. Yet, they yearn to be accepted and integrated into society, disliking being treated differently. Secondly, there is a dichotomy of self-deprecation and self-esteem. Self-deprecation arises from the inner self of the blind, as the loss of sight diminishes their perceived value, leading to feelings of inadequacy. Additionally, societal subconscious biases contribute to the unjust treatment and discrimination of the blind. Due to visual impairment, the blind often impose a strong self-demand for compensating deficiencies, making them more resilient, determined, and self-disciplined than ordinary individuals. Thirdly, there is a lack of a sense of security. Safety needs are fundamental survival elements, second only to physiological needs. Visual impairment weakens the blind's control over themselves and their surroundings, placing them in a cognitively insecure environment. As they struggle to accomplish tasks smoothly, the blind are prone to feelings of frustration and self-doubt. This lack

of a sense of security manifests as suspicion, distrust, and reluctance to engage with others.

Therefore, I aspire to create something that can alleviate their loneliness to a certain extent, fostering social interaction for these "plunged into darkness."

2.2 Chinese Braille

In the upcoming development, I will be using a font known as Chinese Braille. Braille is a writing system specifically designed for the blind, transforming regular text into a touchable and perceptible form using raised or indented dot symbols. Braille has a long history of development, providing the blind with an independent way of reading and writing. Chinese Braille, created by the blind Frenchman Louis Braille in 1824, is different from the international Braille system commonly known as "Braille." The Braille system entered China in 1874, with British missionaries collaborating with Chinese blind individuals. Different dialects led to the creation of various Braille systems, such as the "Kangxi Braille" (the earliest universally used Chinese Braille based on the Kangxi Dictionary), the "Fuzhou Braille" spelling Minnan dialect using Hanyu Pinyin, the "Xinmu Keming Braille" spelling Nanjing Mandarin, and Braille for Cantonese and Hakka dialects. The use of Braille mathematical symbols in China has a history of over ninety years. Since 1911, China has adopted the Taylor symbol system, which continued until the early 1970s when the Marburg symbol system was officially adopted. With the increase in Braille publications and the deepening of content, the Taylor symbol system, only suitable for secondary school-level content, gradually became inadequate[4]. Mr. Huang Jiani, a Braille publisher at the time, conducted in-depth research on the Marburg symbol system and proposed a set of Braille mathematical symbols suitable for China, referencing the Marburg system. In 1990, the China Disabled Persons' Federation and the China Blind Persons' Association jointly held the first expert seminar on Braille mathematical symbols. The experts unanimously agreed that this symbol set, based on the Marburg system and modified and improved by the writing group, was highly systematic, scientific, and practical. This set laid the foundation for the development of higher-level blind education and the publication of higher-level Braille versions of natural science books in China. This is the "Chinese Double Pinyin Braille Scheme" (as shown in Figure 2), the font I will be using.

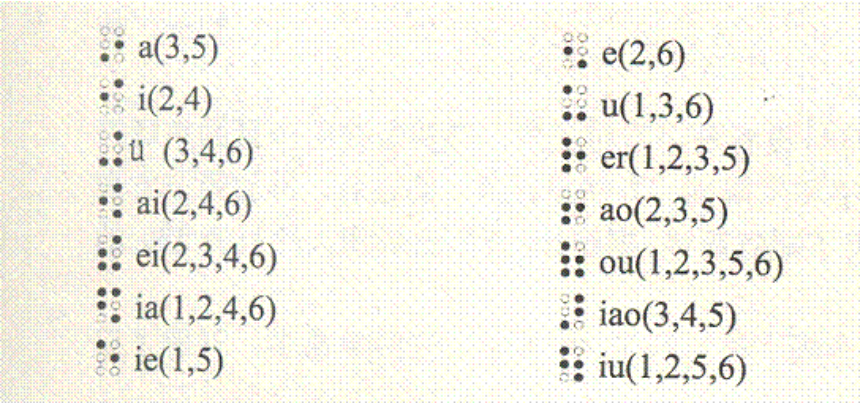

Figure 2

2.3 Deep Learning

Deep learning, akin to the arcane discipline of computers, endows them with an enigmatic wisdom reminiscent of the human brain. These models act as navigators in the cosmic expanse of data, probing intricate patterns from images, text, sound, and other nebulous realms, as if navigating the vast seas of the digital universe rather than frolicking in the shallows of data.The ultimate goal of AI is to cultivate computers with human-like thinking and learning capabilities, as if aiming to turn them into philosophers of the digital age. Deep learning technology serves as the torchbearer in this digital enlightenment, igniting sparks across various applications—from digital assistants and voice-controlled remotes to fraud detection and automatic facial recognition—rendering our daily lives intelligent and entertaining.

Deep learning models, like the celebrities of the computer realm, do not spontaneously emit brilliance; rather, they are the meticulously trained results of data scientists. Using a series of algorithms, much like a masterful chef seasoning a dish, these models are nurtured to handle a myriad of tasks, becoming the multitasking virtuosos of the digital world.

In summary, deep learning empowers computers with a higher level of intelligence, enabling them to interact in a more captivating manner within the digital domain. No longer mere cold processors, they resemble navigators in the starry sky, guiding us to explore novel possibilities in the universe of data.

2.3.1 CNN

Convolutional Neural Networks (CNNs) are a class of feedforward neural networks with convolutional calculations and a deep structure, representing one of the representative algorithms in deep learning. Neural networks mimic the structure and function of the central nervous system, particularly the brain. Neural networks consist of numerous artificial neurons, constructed in different ways to form various networks.

CNNs are one type, alongside others like Generative Adversarial Networks (GANs) and Recurrent Neural Networks (RNNs)[2]. Neural networks can possess simple decision-making and judgment capabilities similar to humans and yield better results in image and speech recognition.

CNNs are widely applied in the field of image recognition. How do CNNs achieve image recognition? The structure can be divided into three layers:

1.Convolutional Layer - Primarily responsible for extracting features.
2.Pooling Layer - Mainly used for downsampling without compromising recognition results.
3.Fully Connected Layer - Primary function is classification[7].

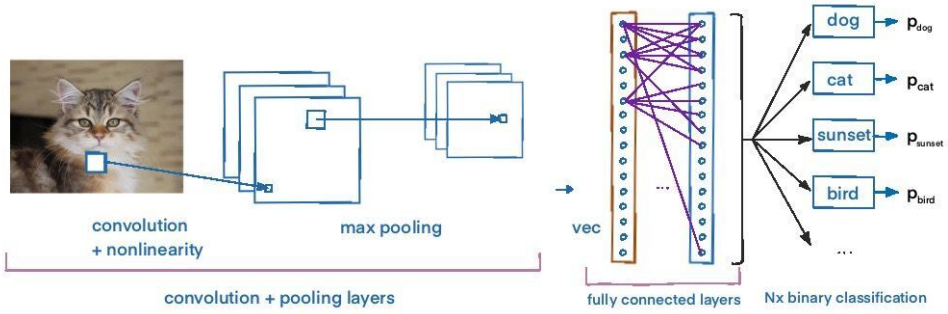

Figure 3

We can draw an analogy between humans and CNNs. For instance, when you see the cat in Figure 3, how does a human recognize it as a cat? First, you assess that the cat's ears are triangular, it has fur all over its body, a soft facial appearance, and a tail. Then, by connecting these features, you deduce that it is a cat. The principle of CNNs is similar: using convolutional layers to identify features and fully connected layers for classification, while pooling layers help reduce the number of parameters and ignore some information while maintaining sampling invariance.

2.4 Related Work

In recent years, research in the field of image recognition has predominantly focused on deep learning technologies, achieving significant advancements. Convolutional Neural Networks (CNNs) have proven highly effective in perceiving the structure of images, automatically extracting distinctive features. However, large neural networks often demand substantial computational power and prolonged training to attain the highest possible accuracy. Addressing the challenge of minimizing the number of parameters in CNNs involves several strategies. Firstly, standardizing the format of training data involves using one-hot encoding in this context. Subsequently, loading the dataset into memory enhances overall performance[8]. Introducing data perturbation ensures that training data is not obtained in the same sequence with each run, while partitioning the data into equally sized batches guarantees consistency in batch sizes

across each epoch. These data manipulation processes are also applied to the testing dataset.

Facial recognition technology involves authenticating individuals by analyzing facial images or videos. With the continuous progress of technology, facial recognition has found widespread applications in areas such as security, finance, and social media. However, due to the complexity and diversity of facial images, achieving high accuracy in facial recognition remains a challenge. To enhance accuracy, deep learning technology is extensively employed.

Deep learning emulates the workings of the human brain's neural network, employing multi-layered neural networks for pattern recognition and feature extraction[11]. In facial recognition, deep learning extracts advanced features from facial images through extensive training data and intricate neural network structures, thereby improving recognition accuracy. Convolutional Neural Networks (CNNs) are utilized for extracting lower-level features, such as edges, textures, and shapes, by employing multiple layers of convolution and pooling operations followed by classification through fully connected layers. Additionally, Recurrent Neural Networks (RNNs) are employed to capture temporal features from facial image sequences, including expressions, postures, and gaze, by utilizing memory units and gate mechanisms to process inputs at different time steps. Furthermore, Generative Adversarial Networks (GANs) are employed to generate more realistic facial images[9]. GANs engage in a competitive interplay between a generator, which produces fake images resembling real ones, and a discriminator, which distinguishes between real and fake images. In facial recognition, GANs learn from a large volume of facial images to generate synthetic images resembling real ones, thereby augmenting training data and improving recognition accuracy.

Despite the significant progress achieved by deep learning technology in facial recognition, challenges persist. Firstly, deep learning technology requires substantial training data and computational resources to train complex neural network models, posing challenges for resource-constrained applications. Secondly, when dealing with complex and diverse facial images, [19]deep learning technology is susceptible to interference from factors such as noise, lighting variations, and occlusions, leading to a decline in recognition accuracy.

To address these challenges, researchers are continually exploring new deep learning technologies and algorithms. Transfer learning enhances facial recognition accuracy by leveraging pre-trained models. Furthermore, improvements and optimizations in Generative Adversarial Networks contribute to enhanced quality and diversity of facial images. Simultaneously, integrating other sensors and technologies, such as infrared cameras and 3D reconstruction, can provide additional information to bolster facial recognition accuracy.

In summary, leveraging deep learning technology can elevate the accuracy of facial recognition. Through Convolutional Neural Networks, Recurrent Neural Networks, and Generative Adversarial Networks, deep learning models extract lower-level features, temporal features, and generate realistic images, achieving more accurate facial recognition. However, ongoing research and refinement of deep learning technology are necessary to address limitations in training data and computational resources, enhancing the robustness and accuracy of facial recognition.

2.5 PAZ

PAZ (Perception for Autonomous Systems) is a hierarchical perception library that provides multiple abstraction levels, empowering users to operate based on their specific needs and skill levels. It consists primarily of three levels: the pipeline, processors, and backend. These abstractions enable users to modularly combine functions in a layered fashion, facilitating preprocessing, data augmentation, prediction, and post-processing of inputs and outputs for machine learning models.By leveraging these abstractions, PAZ constructs reusable training and prediction pipelines for a variety of robotic perception tasks. These tasks encompass 2D keypoint estimation, 2D object detection, 3D keypoint discovery, 6D pose estimation, emotion classification, facial recognition, instance segmentation, and attention mechanisms, among others.

In essence, PAZ is a flexible hierarchical perception library that allows users to customize and combine various functionalities based on their requirements and skill levels. This customization facilitates efficient training and prediction for a multitude of robotic perception tasks.

2.5.1 Enhanced Architectural Design

Within this model, an innovative fully convolutional neural network (CNN) architecture has been meticulously crafted, comprising a strategic ensemble of four residual depthwise separable convolutional layers. Following each convolutional layer, a carefully orchestrated sequence of batch normalization and ReLU activation functions ensues. Culminating in a final layer, this architecture integrates global average pooling and a Softmax activation function, ingeniously producing nuanced predictions. Impressively, this streamlined architecture embodies a lean parameter count, totaling approximately 60,000—a notable tenfold reduction from its progenitor and an astonishing 80-fold decrease from the original CNN model. Despite these parsimonious adjustments, the architecture remarkably attains a formidable 95% accuracy rate in gender classification, a mere 1% diminution from the original implementation. Noteworthy achievements extend to the FER-2013 dataset, where the architecture demonstrates a commendable 66% accuracy rate in emotion classification tasks. Notably, the crowning glory lies in the compact storage of the final architecture's

weights, confined to a modest 855kb file.

This refined architectural strategy not only curtails computational expenses but also empowers researchers to consecutively deploy two models on a single image, seamlessly avoiding pronounced time delays. The holistic pipeline orchestrates OpenCV face detection, gender classification, and emotion classification, exhibiting remarkable efficiency with a runtime ranging from 0.22 to 0.0003 milliseconds on an i5-4210M CPU. This performance translates to a noteworthy 1.5x acceleration when compared to the original architecture.

Augmenting the implementation's sophistication, researchers have introduced a real-time guided backpropagation visualization feature. This augmentation enables a granular examination, revealing which pixels in the image activate elements within higher-level feature maps—a compelling enhancement for insightful analysis and model interpretation.

Chapter 3

Production Process

3.1Emotion Recognition

In the initial stages of the selection of the FER2-2013 dataset, I observed a predominant inclusion of samples featuring European faces. However, in practical applications, particularly when confronted with Chinese faces, I noted a suboptimal accuracy. This discrepancy could potentially be attributed to inherent differences between the samples in the FER2-2013 dataset and Asian faces. To enhance accuracy and cater to Chinese facial features, additional training becomes imperative.

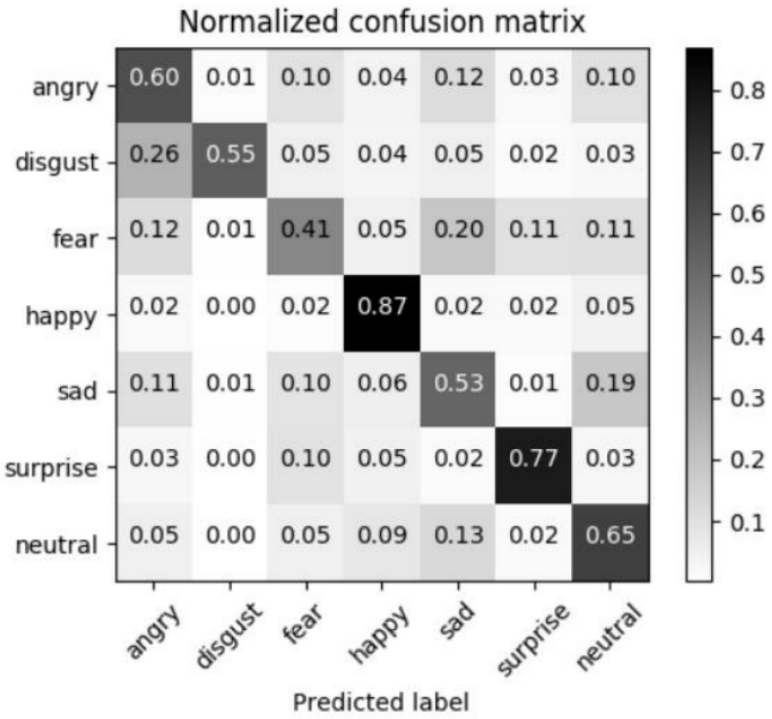

Figure 4

I decided to attempt the collection of more samples featuring Asian faces and blend them with the FER2-2013 dataset. This approach aims to augment the model's cognitive abilities and comprehension specifically regarding Asian facial features. Additionally, the method of transfer learning can be employed. By utilizing a model pre-trained on the FER2-2013 dataset as a foundation and subsequently fine-tuning it on a dataset featuring Asian faces, the convergence speed of the model on Asian faces can be accelerated, thereby improving accuracy.

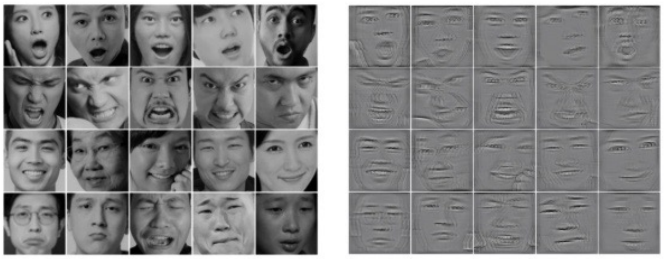

Figure 5

After training, there was a slight improvement in accuracy, reducing the probability of errors in practical scenarios. However, specific environmental requirements are still necessary.

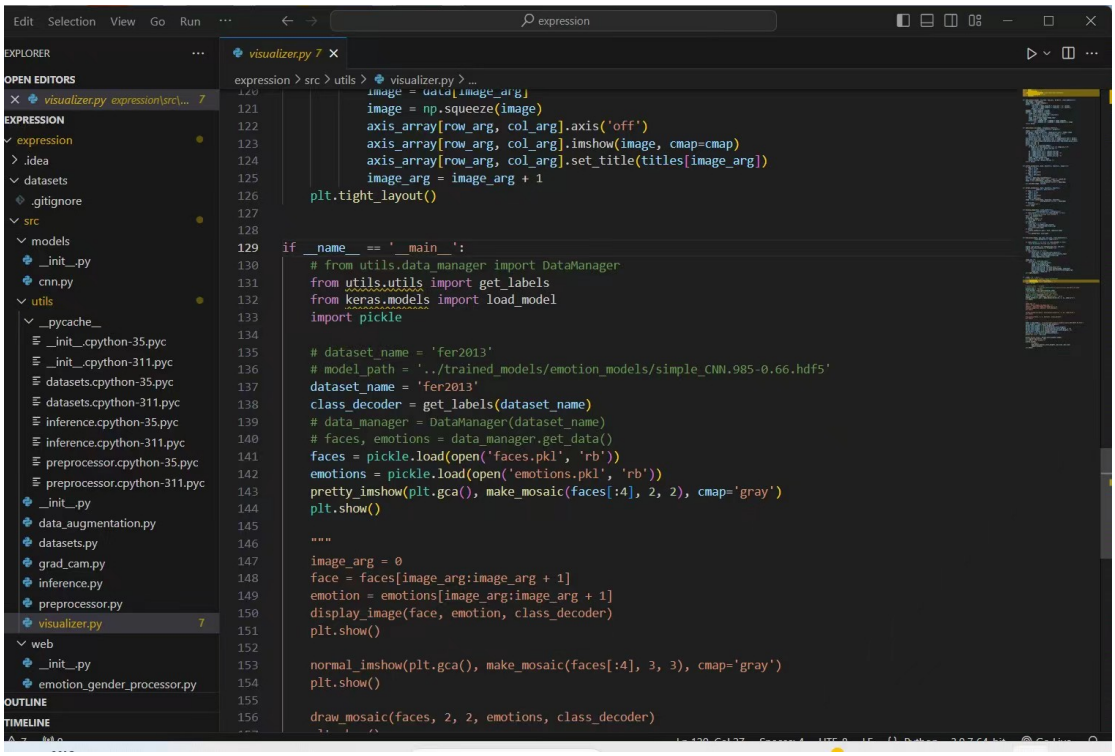

Figure 6

Moreover, if the target wears glasses, it increases the difficulty of recognition. This could be attributed to the identification of the glasses frame as eyebrows during the recognition process, resembling a furrowed brow. Consequently, individuals wearing glasses are more likely to be recognized as expressing anger, regardless of their actual facial expression.

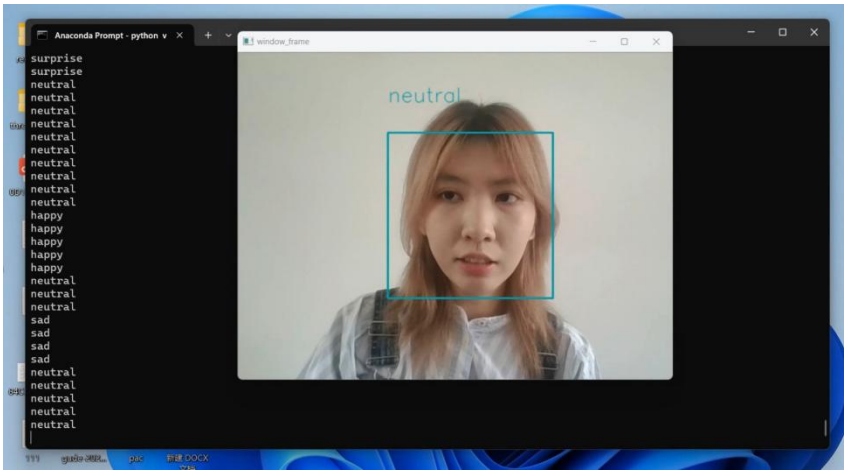

Figure 7

3.2External Arduino Integration

Initially, the decision was made to use the emotions "happy," "natural," "sad," "angry," and "surprise" to create relevant output effects. The goal was to convey the facial expressions of the conversation partner to the user through braille. Among these five

expressions, excluding "natural," the braille corresponding to the other four expressions is illustrated in the following image（figure 8）. As depicted, I needed to create twelve active positions to achieve the desired braille output effect.

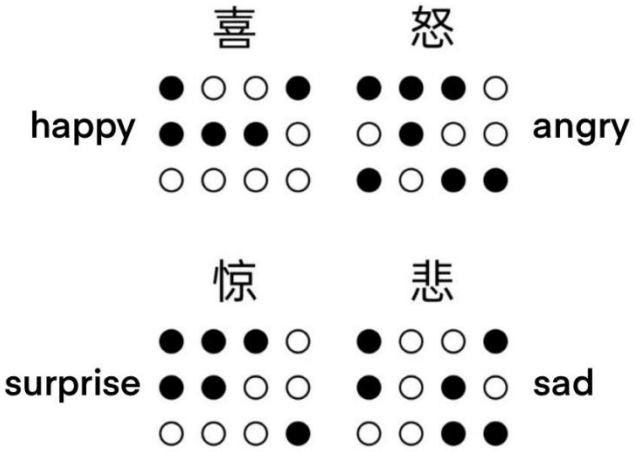

Figure 8

3.2.1Internal Components

Creating an effective tactile interface involves a thoughtful selection of internal components to ensure precision and responsiveness. Here's a detailed breakdown of the key elements:

1.Tactile Modules - Pull-push Solenoids: The backbone of our tactile interface consists of twelve small cylindrical pull-push solenoids. These modules are strategically chosen for their ability to dynamically move tactile points up and down. They play a pivotal role in converting digital signals into tangible, touch-sensitive feedback.

2.Control Center - Arduino Nano ATmega328P Controller: At the core of our system is the Arduino Nano ATmega328P. This compact yet powerful controller processes input signals, orchestrating the nuanced movements of the solenoids. Its adaptability makes it the ideal control center for managing the intricacies of the tactile interface.

3.Power Source - Lithium Battery: Ensuring a consistent power supply, a lithium battery takes center stage. This energy source provides the required voltage to keep the system operational, allowing users to interact seamlessly with the tactile interface.

4.Precision Drivers - L298N Dual-channel DC Motor Driver Boards: Eight L298N dual-channel DC motor driver boards step into the scene to drive the solenoids with precision. These boards play a crucial role in accurately controlling the movement of the tactile points, ensuring a responsive and finely-tuned user interaction.

5.Voltage Management - Voltage Regulator Modules: Two voltage regulator modules

come into play to harmonize the power dynamics. Their role is to adjust the output voltage from the lithium battery, creating an environment conducive to the optimal operation of all system components.

6.Fundamental Elements - Basic Electronic Components: The intricate dance of components is facilitated by fundamental electronic elements such as circuit boards and copper wires. These elements form the connective tissue of the system, ensuring a seamless flow of signals and power.

With these carefully selected and orchestrated internal components, the design and construction of the circuit diagram (refer to Figure 9) become the blueprint for an interactive tactile interface. The Arduino Nano, acting as the conductor, orchestrates the tactile symphony, turning digital signals into a tangible and interactive experience for users with visual impairments.

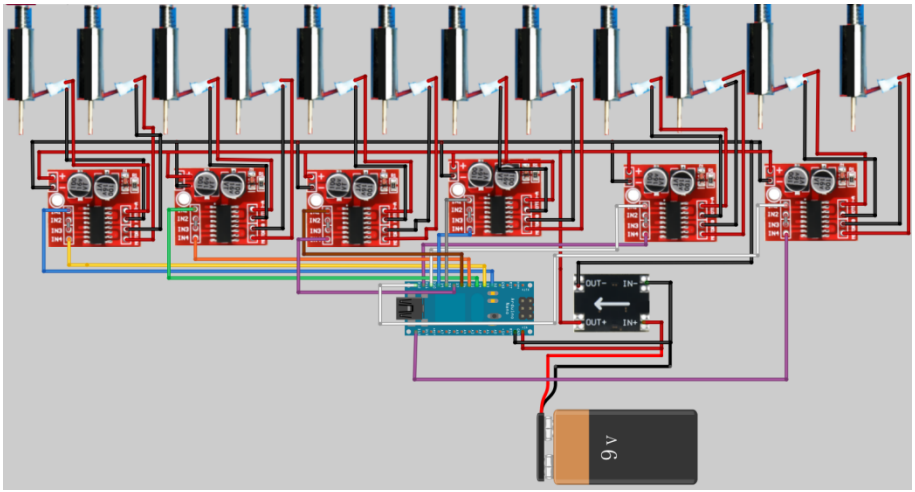

Figure 9

With the selection of components laid out, the next phase involves the meticulous assembly of these elements to create a cohesive and functional tactile interface. The assembly process is a symphony of precision and care, ensuring each part plays its role seamlessly.

1.Strategic Placement of Solenoids: The heart of the tactile interface lies in the placement of the pull-push solenoids. These components are strategically positioned to achieve optimal tactile feedback. Precision in their arrangement is key to creating a nuanced and effective touch-sensitive surface.

2.Arduino Integration: The Arduino Nano ATmega328P takes its place, integrated with the solenoids through the L298N dual-channel DC motor driver boards. This integration forms the nerve center of the system, where digital commands are translated into tangible responses.

3.Battery Power Integration: Careful attention is given to incorporating the lithium battery into the system. The power source is seamlessly integrated, ensuring a stable

and reliable energy supply to sustain the tactile interface's continuous operation.

4.Fine-tuning with L298N Boards: The eight L298N dual-channel DC motor driver boards are delicately tuned to synchronize with the solenoids. This step is crucial for achieving the desired precision in controlling the movement of the tactile points, providing users with a responsive and accurate interaction experience.

5.Voltage Regulation Refinement: The voltage regulator modules play a vital role in fine-tuning the power dynamics. Their integration ensures that each component receives the optimal voltage, contributing to the overall efficiency and longevity of the tactile interface.

Circuitry and Wiring Choreography: Basic electronic components, such as circuit boards and copper wires, are intricately woven into the assembly process. The choreography of these elements is meticulously planned to facilitate seamless communication and power distribution among all parts.

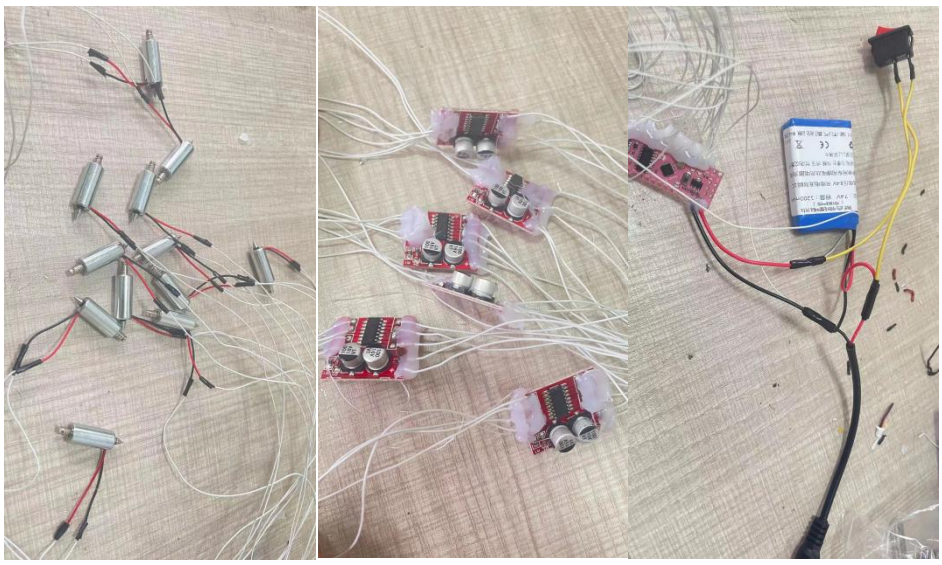

Figure 10

3.2.2Crafting the Exterior

The next phase of the project involves translating conceptualizations into tangible form—an enclosure that not only accommodates but also enhances the functionality of the twelve small cylindrical pull-push solenoids. The sketch provided (see Figure 11) serves as a blueprint for this purpose, detailing a four-part structure comprising a lid, body, base, and twelve interactive buttons.The lid, akin to a protective canopy, is meticulously crafted to shield the intricate components within. Its design not only ensures the safety of the internal mechanisms but also provides convenient access for maintenance when required.The body of the enclosure serves as the primary housing

unit, carefully designed to accommodate the solenoids, Arduino Nano, L298N driver boards, and other essential components. The spatial arrangement within the body is optimized for efficient use of space, fostering a harmonious coexistence of elements.

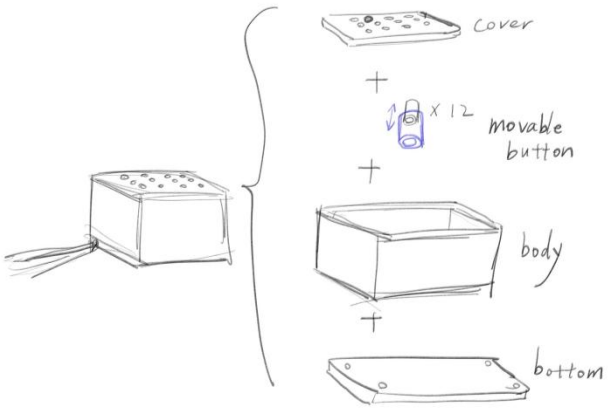

Figure 11

Embarking on the modeling phase using SolidWorks, I translated the conceptual sketch into a tangible 3D model. This digital transformation involved merging the lid and body of the box, concurrently securing the lower portions of the interactive buttons. The SolidWorks model, illustrated in Figure 11， 12， 13 below, captures the essence of the envisioned enclosure.

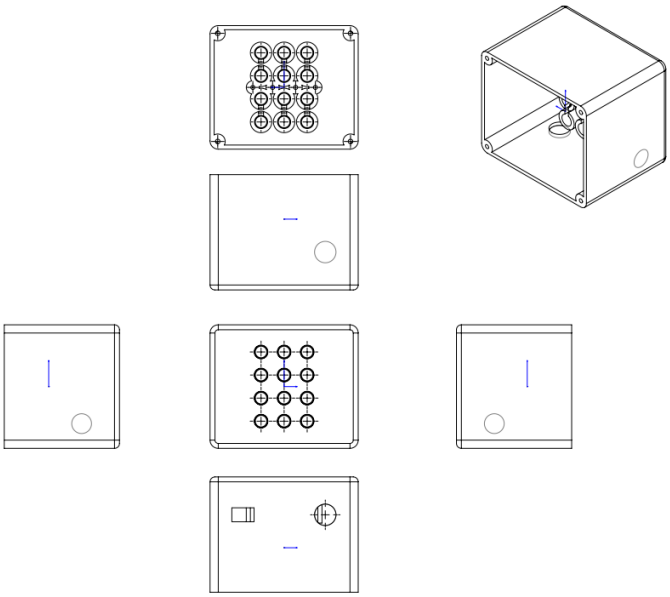

Figure 12

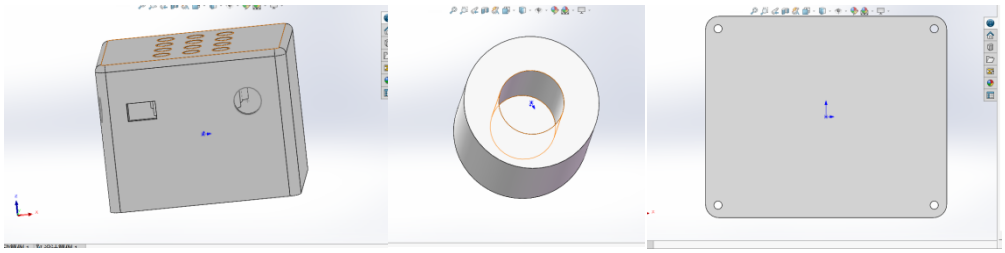

Figure 13

With the SolidWorks model finalized, the next pivotal step is the realization of the digital design into a physical form through 3D printing. The chosen material, Polyamide, offers a balance of durability and flexibility, aligning seamlessly with the functional requirements of the envisioned enclosure.The utilization of Polyamide in the 3D printing process brings the meticulously crafted SolidWorks model into the physical realm. Layer by layer, the printer deposits Polyamide material, gradually building the enclosure according to the specifications encoded in the digital design. This additive manufacturing approach ensures precision and accuracy in material deposition.Polyamide, known for its robustness and flexibility, lends itself well to the intended functionality of the enclosure. The material's properties provide the necessary structural integrity to protect the internal components while allowing sufficient flexibility for the interactive buttons' responsive movements.（figure 14，15）

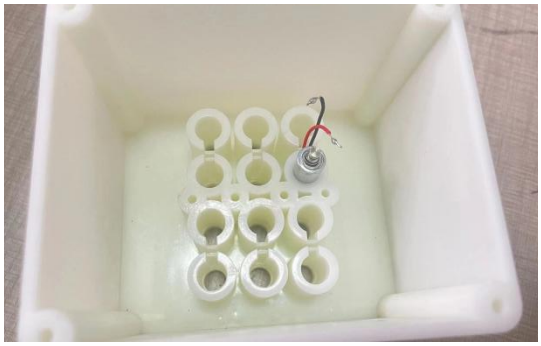

Figure 14

Figure 15

the next pivotal phase involves the development and fine-tuning of the underlying

code that governs the behavior of the interactive system.

```
    if len(emotion_window) > frame_window:
        emotion_window.pop(0)
    try:
        emotion_mode = mode(emotion_window)
    except:
        continue

    if emotion_text == 'angry':
        color = emotion_probability * np.asarray((255, 0, 0))
        time.sleep(1)
        ser.write(b'a')
    elif emotion_text == 'sad':
        color = emotion_probability * np.asarray((0, 0, 255))
        time.sleep(1)
        ser.write(b's')
    elif emotion_text == 'happy':
        color = emotion_probability * np.asarray((255, 255, 0))
        time.sleep(1)
        ser.write(b'h')
    elif emotion_text == 'surprise':
        color = emotion_probability * np.asarray((0, 255, 255))
        time.sleep(1)
        ser.write(b'x')
    elif emotion_text == 'neutral':
        color = emotion_probability * np.asarray((0, 255, 255))
        time.sleep(1)
        ser.write(b'o')
    else:
        color = emotion_probability * np.asarray((0, 255, 0))
```

Figure 16

```
expression > src > models > cnn.py > mini_XCEPTION > output
253
254     x = MaxPooling2D((3, 3), strides=(2, 2), padding='same')(x)
255     x = layers.add([x, residual])
256
257     # module 3
258     residual = Conv2D(64, (1, 1), strides=(2, 2),
259                       padding='same', use_bias=False)(x)
260     residual = BatchNormalization()(residual)
261
262     x = SeparableConv2D(64, (3, 3), padding='same',
263                         kernel_regularizer=regularization,
264                         use_bias=False)(x)
265     x = BatchNormalization()(x)
266     x = Activation('relu')(x)
267     x = SeparableConv2D(64, (3, 3), padding='same',
268                         kernel_regularizer=regularization,
269                         use_bias=False)(x)
270     x = BatchNormalization()(x)
271
272     x = MaxPooling2D((3, 3), strides=(2, 2), padding='same')(x)
273     x = layers.add([x, residual])
274
275     # module 4
276     residual = Conv2D(128, (1, 1), strides=(2, 2),
277                       padding='same', use_bias=False)(x)
278     residual = BatchNormalization()(residual)
279
280     x = SeparableConv2D(128, (3, 3), padding='same',
281                         kernel_regularizer=regularization,
282                         use_bias=False)(x)
283     x = BatchNormalization()(x)
284     x = Activation('relu')(x)
285     x = SeparableConv2D(128, (3, 3), padding='same',
286                         kernel_regularizer=regularization,
287                         use_bias=False)(x)
288     x = BatchNormalization()(x)
```

Figure 17

After assembling the device, I moved on to the crucial step of debugging the code, focusing on four specific emotions: anger, sadness, joy, and surprise. To begin, I connected the device and launched the Arduino IDE. I made sure to open the correct port and, if needed, adjusted the port settings. Alternatively, I initiated the terminal to activate my white box capable of emotion recognition.

Chapter 4

Artwork

The final implementation involves the listener placing their hand on the white box, while the person being observed opens the camera window. Through the camera, the emotions of the observed individual are recognized. The Python script then sends this data to the Arduino, which activates the tactile buttons on the white box to represent the detected emotion. There are five different scenarios:

1. When the recognized emotion is happy, the results in the window and the status of the white box's buttons are as shown in Figures 18 and 19:

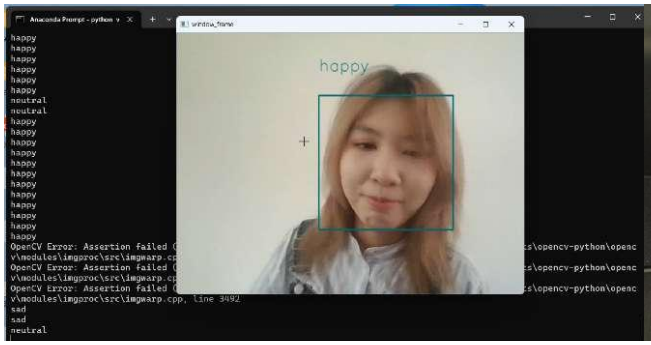

Figure 18

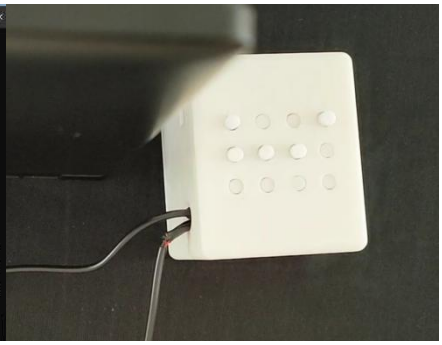

Figure 19

2. When the recognized emotion is sad, the results in the window and the status of the white box's buttons are as shown in Figures 20 and 21:

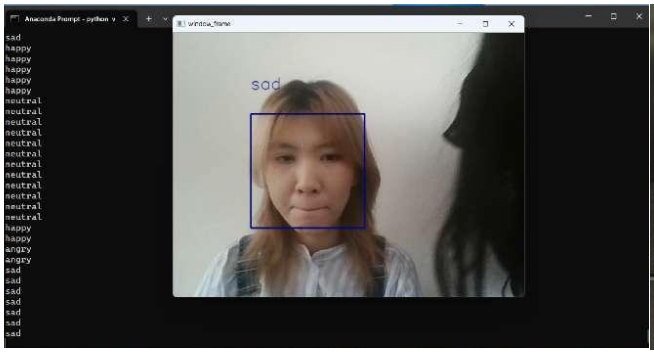

Figure 20

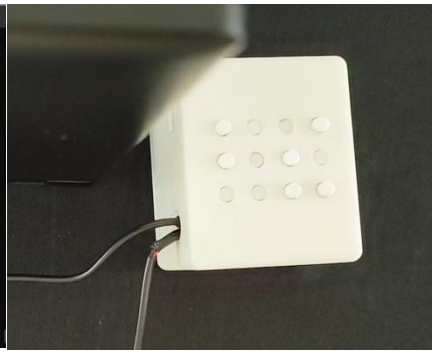

Figure 21

3. When the recognized emotion is anger, the results in the window and the status of the white box's buttons are as shown in Figures 22 and 23:

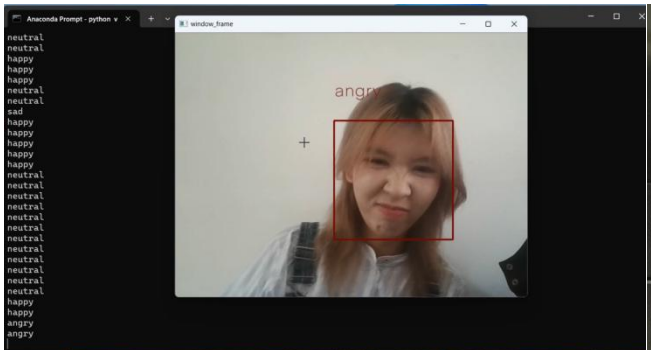

Figure 22

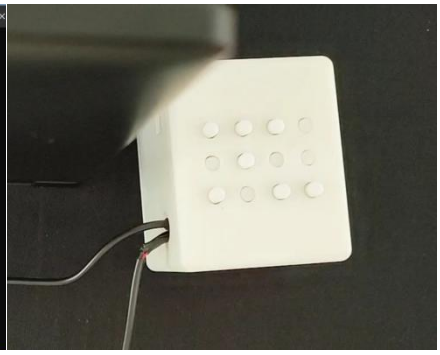

Figure 23

4. When the recognized emotion is surprise, the results in the window and the status of the white box's buttons are as shown in Figures 24 and 25:

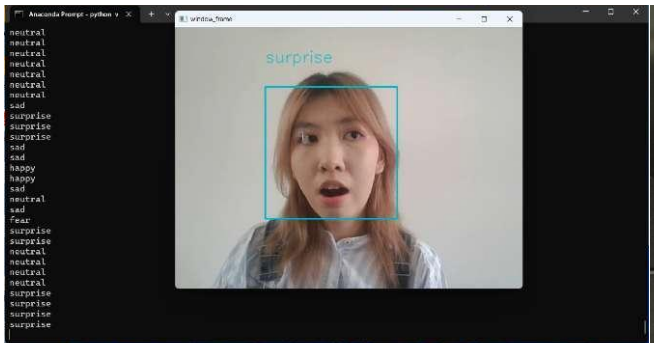

Figure 24

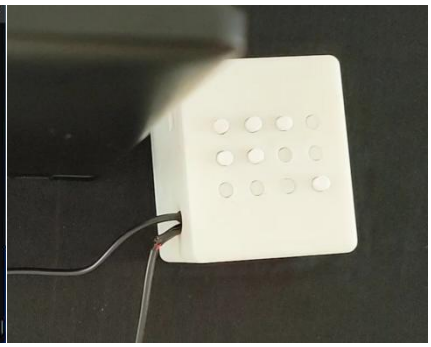

Figure 25

5. When the recognized emotion is natural, the results in the window and the status of the white box's buttons are as shown in Figures 26 and 27:

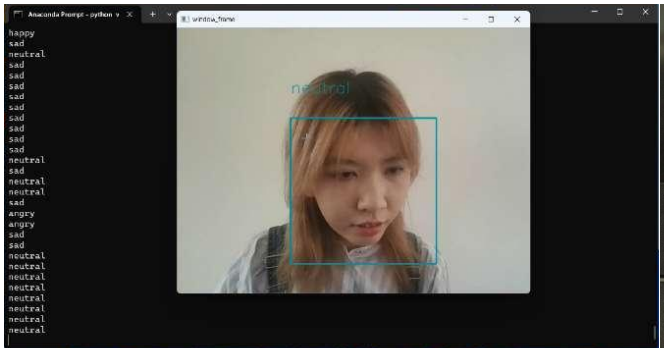

Figure 26

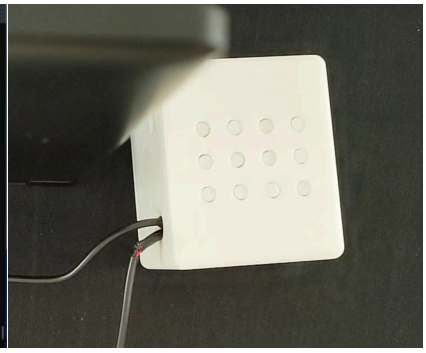

Figure 27

Additionally, even with slight facial obstructions, accurate detection of facial expressions in front of the camera is maintained, as shown in Figure 28:

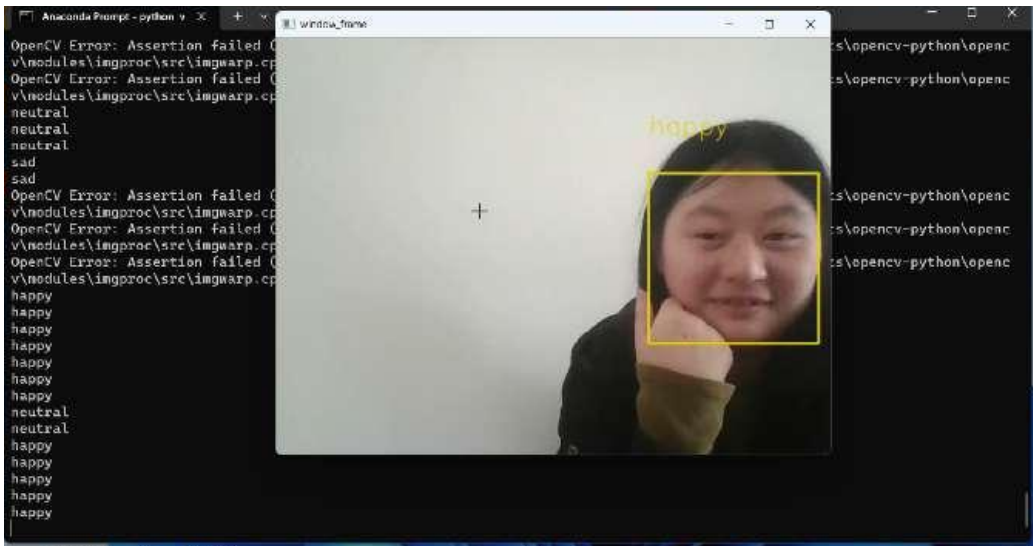

Figure 28

After collecting more samples of Asian faces and merging them with the FER2-2013 dataset to enhance the model's recognition of Asian faces, there was a slight improvement in emotion detection accuracy. In practical applications, errors in judging Asian facial expressions were reduced. Initially, there were voltage issues with the Arduino components, but after adjustments and utilizing an adjustable power supply, the problem was resolved, and the system ran smoothly.

Chapter 5

Evaluation

In China, I engaged eight volunteers divided into four pairs, each consisting of a blind listener and a observed individual, for a unique experiment designed to assess the performance of the project in situations where visibility is limited.

5.1 Personal Information

Among these four pairs of volunteers, two were completely blind, one had cataracts and could only see object outlines, and one had normal vision but experienced simulated blindness by covering their eyes.

Name	Lu Xiaoyan	Wang Jun	Dong hui	Wang qinye
Age	19	24	41	50
Gender	female	male	male	female
Condition	healthy	cataract	blind	blind

Table 1: Personal information of the listener

Name	Shen Tong	Wu Yongfan	Huang Jiashan	Li Yunsheng
Age	23	52	20	53
Gender	female	female	male	male
Condition	healthy	healthy	healthy	healthy

Table 2: Personal information of the talker

5.2 Questionnaire

During the experiment, the blind listener placed their hand on a white box to sense the emotions of the observed individual, rather than directly observing their facial expressions. This setup prevented the blind listener from visually assessing the emotional state of the observed individual, making them rely more on the performance of the project. In the experiment, the observed individual needed to display various emotions while engaging in a natural conversation with the blind listener, testing how well the project performed in this specific scenario. During the conversation, the blind listener had to understand and respond to the emotional states conveyed by the project. This experimental design aimed to simulate real-life communication scenarios to assess the effectiveness and accuracy of the project in practical applications. Through this experiment, we could gain a deeper understanding of the project's applicability in different situations, especially for individuals with visual impairments. It also helped identify potential issues in real-world applications, providing valuable insights for future improvements. Such endeavors contribute not only to refining the project itself but also to advancing related technologies, delivering practical value to a broader user base.

Question 1: Have you encountered this type of product before?

Name	Lu Xiaoyan	Wang Jun	Dong hui	Wang qinye
Yes or No	No	No	No	No

Table 3: information of the listener

Name	Shen Tong	Wu Yongfan	Huang Jiashan	Li Yunsheng
Yes or No	No	No	No	No

Table 4: information of the talker

Question Two: Have you encountered Braille before?

Name	Lu Xiaoyan	Wang Jun	Dong hui	Wang qinye
Yes or No	No	Yes	Yes	Yes

Table 5: information of the listener

Name	Shen Tong	Wu Yongfan	Huang Jiashan	Li Yunsheng

Yes or No	No	Yes	No	Yes

Table 6: information of the talker

Question Three: Is it easy to operate?

Name	Lu Xiaoyan	Wang Jun	Dong hui	Wang qinye
Yes or No	Yes	Yes	Yes	Yes

Table 7: information of the listener

Name	Shen Tong	Wu Yongfan	Huang Jiashan	Li Yunsheng
Yes or No	Yes	No	No	Yes

Table 8: information of the talker

Question Four: How does the touch feel of the white box?

1.Lu Xiaoyan: Smooth, no sharp edges.
2.Wang Jun: Gets a bit warm after holding it for a while.
3.Dong Hui: Smooth, becomes warm.
4.Wang Qinye: It's okay.

Question Five: Can you smoothly receive the emotions expressed by the white box?

1.Lu Xiaoyan: Need to memorize the patterns for a while, otherwise can't recognize.
2.Wang Jun: Generally can.
3.Dong Hui: Can feel it.
4.Wang Qinye: Can feel it, but it's a bit large.

Question Six: How do you feel about the accuracy of expressing emotions?

1.Shen Tong: Sometimes it's strange.
2.Wu Yongfan: Seems okay.
3.Huang Jiashan: Feels like there are many mistakes.
4.Li Yunsheng: Mostly correct.

Question Seven: Does the accuracy of expressing emotions have any impact on the conversation?

1.Lu Xiaoyan: Seems to have no impact.
2.Wang Jun: Not really.
3.Dong Hui: Sometimes don't know whether to cry or laugh.
4.Wang Qinye: Still hope it's all correct.
Talker:
1.Shen Tong: Doesn't seem to have much impact, probably.
2.Wu Yongfan: Some impact.

3.Huang Jiashan: Has an impact, feels like emotions are not conveyed.
4.Li Yunsheng: Still not used to it.

Question Eight: Does the sound or touch of the work have any impact on the conversation?

1.Lu Xiaoyan: No sound.
2.Wang Jun: No significant impact.
3.Dong Hui: Notices the mechanical sound of the buttons.
4.Wang Qinye: There's sound but it doesn't affect the conversation.

Question Nine: How is the overall user experience?
Listener:

1.Lu Xiaoyan: Memorizing braille is a bit troublesome, but feels magical after remembering.
2.Wang Jun: Initially couldn't figure out two expressions, took a couple of tries.
3.Dong Hui: Something never seen before, but it works.
4.Wang Qinye: A strange feeling, as if you can really see the expressions.
Talker:
1.Shen Tong: It's interesting to see your own recognized expressions.
2.Wu Yongfan: Quite interesting.
3.Huang Jiashan: Feels like it can be improved.
4.Li Yunsheng: She thinks it's good as long as she likes it.

5.3Summary

This project demonstrates excellent accuracy in facial expression recognition in real-world scenarios. For first-time users, it undoubtedly presents an attractive and innovative technology. The project effectively conveys various emotions, and users can easily understand and receive the emotional information it communicates. In terms of operation, the project is quite user-friendly, with minimal noise and smooth tactile feedback. For both parties engaged in conversation, the entire process is not heavily disrupted.

However, despite the project's ability to accurately recognize expressions in most cases, it has certain limitations in capturing subtle facial expressions. When the observed individual rapidly switches between multiple expressions in a short period, the project may not immediately capture these changes and accurately express various emotions. This limitation may, to some extent, affect the effectiveness of emotional communication.

Furthermore, due to the design of the project, both the listener and the observed

individual may experience some posture restrictions during use. If a person moves out of the camera's field of view, the emotion recognition function will not work correctly, leading to a failure in emotional transmission. Therefore, users need to be aware of these limitations when using the project and operate it in suitable environments and conditions to ensure the best communication experience.

5.4 Discussion

In interpersonal communication, facial expressions play a crucial role as a non-verbal communication method. However, for individuals who are visually impaired and unable to perceive others' expressions visually, they lose a vital means of communication with the outside world. Being visually impaired makes them more susceptible to loneliness and depression than healthy individuals[6]. This project, represented by a small white box, aims to take a small step forward for them, drawing more attention and care from healthy individuals. In reality, losing vision significantly increases the difficulty of using electronic devices for visually impaired individuals, let alone understanding or learning code. However, in today's digital age, with the increasing prevalence of AI and the widespread use of ML, these technologies are bound to become accessible to them sooner or later.

Personal Evaluation

Overall, I am quite satisfied with the current state of this project. Throughout this endeavor, I successfully employed tools such as OpenCV and Convolutional Neural Networks (CNN) to achieve facial emotion recognition and seamlessly integrated external connectivity with Arduino. I managed to recruit volunteers to test the "white box" and provide valuable feedback. Additionally, I conducted relevant training for the model, enhancing its accuracy in recognizing facial expressions in Asian faces. Engaging in conversations with volunteers revealed insights that I wouldn't typically notice, enriching my understanding of the project. Although the process of modifying the dataset was challenging, the sense of accomplishment I gained from successfully loading new datasets and adjusting images and data multiple times was significant.

Chapter 6

Conclusion

6.1Future Work

While the Little White Box has made significant strides in providing utility, there are still many areas for improvement. For instance, the accuracy of facial recognition can

be enhanced through additional training on diverse datasets or further fine-tuning of the model. Additionally, the design of the Little White Box can be optimized, exploring the possibility of a more compact casing and implementing measures to minimize heat generation. These improvements aim to enhance the performance and user experience of the Little White Box, making it more effective in catering to the emotional communication needs of visually impaired individuals. Future work will focus on research and enhancements in these areas to continually refine the application of this innovative technology.

Bibliography

[1]. Author(s) (2023). Individual differences in spontaneous facial expressions in people with visual impairment and blindness. British Journal of Visual Impairment, (3), 475-488.

[2]. Author(s) (2022). CNN-based efficient approach for emotion recognition. Journal of King Saud University - Computer and Information Sciences, (9), 7335-7346.

[3]. Yang, J., Qiao, P., Li, Y., & Wang, N. (2019). A comprehensive review of machine learning classification problems and algorithms. Statistics and Decision, (06), 36-40. doi:10.13546/j.cnki.tjyjc.2019.06.008.

[4]. Zhang, W. (2018). A review and reflection on the history of braille publishing in China. Research on the History of Publishing in China, (04), 33-49. doi:10.19325/j.cnki.10-1176/g2.2018.04.005.

[5]. Author(s) (2018). Hallucinations in an Elderly Patient with Severe Visual Impairment. Cureus, (11), e3592.

[6]. Author(s) (2018). The everyday lives of older adults with visual impairment: An occupational perspective. British Journal of Occupational Therapy, (5), 266-275.

[7]. Meng, D. (2017). A research overview of image classification methods based on deep learning. [Title of Dissertation]. Retrieved from [Link to dissertation].

[8]. Author(s) (2016). Convolutional neural networks in image understanding. Acta Automatica Sinica, (09), 1300-1312. doi:10.16383/j.aas.2016.c150800.

[9]. Lu, G., He, J., Yan, J., & Li, H. (2016). A convolutional neural network for facial expression recognition. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), (01), 16-22. doi:10.14132/j.cnki.1673-5439.2016.01.003.

[10]. Zheng, Y., Quan, R., & Zhang, Y. (2014). New developments in deep learning and its applications in target and behavior recognition. Journal of Image and Graphics, (02), 175-184.

[11]. Chen, X. (2014). Research on deep learning algorithms and applications based on convolutional neural networks. [Title of Dissertation]. Retrieved from [Link to dissertation].

[12]. Author(s) (2013). A survey of the use of braille in China. Language and Word Application, (02), 42-48. doi:10.16499/j.cnki.1003-5397.2013.02.002.

[13]. Cai, R. (2012). Principle and application of Arduino. Electronic Design Engineering, (16), 155-157. doi:10.14022/j.cnki.dzsjgc.2012.16.031.

[14]. Author(s) (2012). Research overview of deep learning. Computer Applications Research, (08), 2806-2810.

[15]. Zou, L., & Zhang, X. (2012). Artificial intelligence and its development applications. Information Network Security, (02), 11-13.

[16]. Yang, C., & Che, L. (2011). Design of Chinese Braille conversion system. Journal of Beijing Institute of Graphic Communication, (06), 36-38. doi:10.19461/j.cnki.1004-8626.2011.06.009.

[17]. Zhu, J. (2009). Application and research of SolidWorks software in mechanical design. New Technology and New Process, (02), 41-44.

[18]. Hui (2007). A brief history of China Braille Publishing House. Publishing History Materials, (04), 13.

[19]. Liu, X., Tan, H., & Zhang, Y. (2006). New developments in facial expression recognition research. Journal of Image and Graphics of China, (10), 1359-1368.

[20]. Author(s) (1998). [Title of the Article]. China Disabled Persons, (10), 6-7.

[21]. Gao, W., & Jin, H. (1997). Analysis and recognition of facial expression images. Journal of Computer Science, (09), 782-789.

[22]. Singh, N.K. and Pal, N.R. (2023) Convolutional neural networks exploiting attributes of biological neurons, arXiv.org. Available at: https://arxiv.org/abs/2311.08314 (Accessed: 24 November 2023).