Utilizing LSTM Models for IMU-Based Hand Gesture Classification

Guotao Gou, Yiheng Zhang, Chenhan Zhang, Ziwei Huang, Dekai Feng, Yuanzhuo Zhang

***Abstract*—This course project explores the application of Inertial Measurement Unit (IMU) technology in predicting and classifying multiple hand gestures, with a focus on enhancing the user experience for wearable robotic devices. By employing advanced signal processing and machine learning techniques, we have developed a system capable of accurately interpreting various grasping intentions. The project leverages an LSTM (Long Short-Term Memory) model to process IMU data collected from participants, achieving a high level of accuracy in gesture recognition. The outcomes of this project are expected to contribute to the development of more intuitive and responsive wearable technologies, improving the interaction between users and their devices.**

***Keywords*—Intent Recognition, Long Short-term Memory (LSTM), Inertial Measurement Unit (IMU).**

I. Introduction

In the realm of human-computer interaction, the ability to control devices with natural hand gestures has the potential to revolutionize the way we engage with technology. The advent of wearable robotic devices has brought this possibility closer to reality, yet the challenge remains to create a system that can accurately interpret the user's intentions from their hand movements. This course project aims to address this challenge by utilizing IMU technology to predict and classify a range of hand gestures [1], [2].

The project's foundation lies in the IM900 IMU, a state-of-the-art sensing device that provides rich motion data with high temporal resolution. By strategically positioning the IMU on the forearm, we ensure that the collected data accurately reflects the user's arm movements without interference from hand motion. The data is then processed to calculate the trajectory and orientation of the arm, which are critical features for gesture classification.

Our approach to gesture recognition involves training an LSTM model, a type of deep learning model particularly suited for handling sequential data. The model is trained on a dataset where each gesture is labeled and represented by its corresponding arm movement features. Through rigorous testing and validation, the LSTM model has demonstrated a high degree of accuracy in classifying the gestures, even when faced with new, unseen data [3], [4]

The success of this project holds significant implications for the future of wearable technology. By improving the accuracy and responsiveness of gesture recognition, we can create devices that are more intuitive and user-friendly. This not only enhances the user experience but also opens up new possibilities for applications in various fields, from gaming and virtual reality to assistive technologies for individuals with motor impairments.

After detailed market research, we have determined to use this technology to assist people with disabilities in their daily lives, restoring their ability to survive, live, and even work as much as possible. According to statistics, there were approximately 1.3 billion people with disabilities worldwide in 2023, and over 85 million people with disabilities in China in 2024. However, there are few mature products that can assist people with disabilities through intention recognition methods. This IMU based intention recognition technology has advantages such as accuracy, safety, simplicity, and low cost, and will be a blessing for people with disabilities in China and even the world. Based on market demand analysis and literature research, it is reasonably predicted that if this technology can be commercialized, it will have extremely high market recognition.

At the same time, this technology can promote social development. Since the 18th CPC National Congress, the CPC Central Committee, with Comrade Xi Jinping at its core, has shown "extra care and concern" for the cause of persons with disabilities, standing on the height of seeking happiness for the Chinese people and rejuvenation of the Chinese nation, safeguarding the basic livelihood of persons with disabilities, improving the quality of life of persons with disabilities, and promoting the comprehensive development of persons with disabilities, so that 85 million persons with disabilities have become equal members of large families in the society. The 85 million persons with disabilities have become equal members of the extended family and are sharing the great glory of the country's prosperity and strength with the people of China.

From this, it can be seen that the development, application, and popularization of this technology will have a huge impact on the realization of human rights and the improvement of social morality in the entire society. We can expect that this will have profound significance for the country and even in reality.

In summary, in the field of technology, this course project represents a step forward in the field of human-computer interaction, demonstrating the potential of IMU technology and machine learning in creating more natural and efficient device interaction methods for users. Meanwhile, it has extremely high market value and social significance, making it a technology worthy of in-depth research. [5], [6].

II. Related works

In addition to IMU based intent recognition technology and products, there are also some other technologies and products that have achieved intent recognition in the market.

The following are several main non IMU intent recognition technologies:

1. Computer vision technology: This type of technology utilizes cameras and image processing algorithms to identify user intentions. For example, Microsoft's Kinect system achieves gesture control and intention recognition by capturing and analyzing user body movements [7]. This type of technology is widely used in games, virtual reality, and gesture control devices.

2. Brain Computer Interface (BCI): BCI technology identifies the user's intention by recording and analyzing EEG waves. Users control computers or other devices through their minds, and this technology has important applications in medical rehabilitation and assistive devices. For example, head worn devices provided by companies such as Neurable and Emotiv [8], [9].can achieve intention recognition based on brainwaves.

3. Electromyography (EMG) sensor: The EMG sensor infers user intention by detecting muscle activity electrical signals. This type of technology plays an important role in prosthetic control and motor rehabilitation. For example, Myo Armband is an EMG based arm strap that can be used for gesture recognition and device control [10].

4. Speech recognition: Speech recognition technology identifies a user's intention by analyzing their speech instructions. This technology is widely used in smart assistants (such as Amazon Alexa, Google Assistant [11], [12]) and voice control devices. Through natural language processing (NLP) and deep learning algorithms, the system is able to understand and execute user voice commands

TABLE I
CAPTION

III. Data

A. calibration

Before the formal data acquisition, we successively used the acceleration calibration method to place the IMU on its six faces successively to ensure the consistent acceleration under various attitudes. With the magnetometer calibration method, we rotated the IMU on the horizontal and vertical axes and corrected the IMU measurement data through the magnetic field change to eliminate the initial starting error

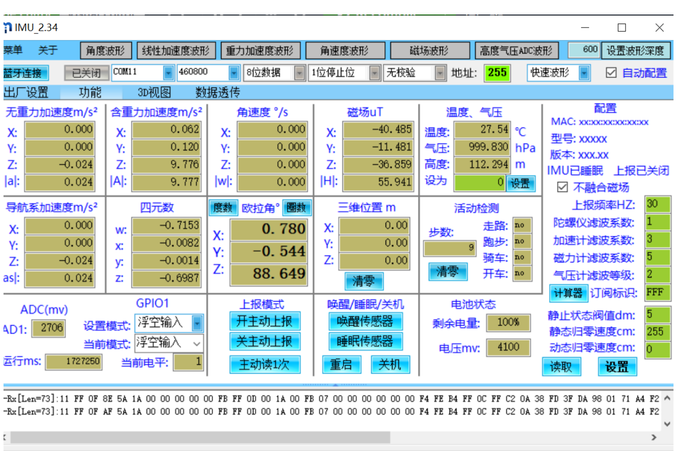

Fig. 1. Software interface for collecting IMU data

B. offline

In this part, we use the software IMU_2.34 developed by IM900 manufacturer Chenyi Electronic Technology for the IMU sensor to collect offline data. It includes the acceleration under the sensor reference frame (with and without gravitational acceleration), the acceleration under the Earth reference frame (without gravitational acceleration), angular velocity, quaternion, angular position, translational position and magnetic field, temperature and other data.

In the process of offline data acquisition, after opening the record, we naturally and smoothly draw the Arabic numerals "1", "2", "3" in the spatial vertical plane with the amplitude of the elbow joint of the main moving human body and the shoulder joint of the small moving human body within about two seconds. After each number is crossed, the data recording is stopped and stored in the offline database

C. online

In this part, through BLE5.0 Bluetooth communication protocol, the data body containing the function label and content is directly transferred from the IMU hardware to the computer storage area, and is retrieved in real time during the machine learning process

In the online data collection process, for the consideration of simple, fast and easy popularization of hand gestures, we changed the collected hand gestures from Arabic numerals "1", "2", "3" to four simple strokes up, down, left and right in the vertical plane of space. This time, for the consideration of universality, in the process of motion data collection, when we deliberately changed the posture of the waist joint of the human body, the hand movement included four directions of motion mainly based on shoulder joint movement, and some movements mainly based on elbow joint movement, and there was also a combination of the two.

D. data detail

Considering the working principle of IMU, direct measurement of acceleration data has the highest accuracy. In the offline data learning stage, we use the IMU_2.34 software to record aX(m/s²), aY(m/s²) and aZ(m/s²) in the file. That is, the acceleration of IMU hardware reference frame after removing the gravitational acceleration is used as training and test data.

However, in the online data collection process, we found that aX(m/s²), aY(m/s²) and aZ(m/s²) were not effective in learning the three data. Considering that online learning has higher requirements for computing speed, we successively switched to aX(m/s²), aY(m/s²), Az(m/s²), Ax(m/s²), Ay(m/s²) and AZ (M /s²). In other words, IMU hardware reference acceleration without gravity acceleration and asX(m/s²), asY(m/s²), asZ(m/s²), namely navigation system acceleration without gravity acceleration, are used as learning data. Finally, the latter is adopted from the perspective of learning success rate

IV. Material and Methods

A. Intent Recognition

1) Data Processing: The absolute acceleration data obtained from the IMU served as a reference for determining the initiation of arm movement, based on an analysis of experimental data. We established a threshold where an absolute acceleration greater than 0.5 m/s^2 indicated the start of movement, marking this as the starting point. Once movement commenced, we calculated the velocity in the x, y, and z directions by integrating the acceleration data from the starting point. A subsequent integration of these velocities provided us with the distances traveled in each direction, representing the trajectory of the arm. Similarly, by integrating the angular velocities in the x, y, and z directions from the starting point, we obtained the angles in each axis, representing the orientation of the arm. Such processing of the input data imbues it with practical physical significance, enhancing the accuracy of classification and making the data more comprehensible.

2) Deep Learning Process: We obtain the distances and angles relative to the starting point in the x, y, and z axes for each sample. These represent the trajectory and orientation of the arm, respectively. And the data are padded or truncated to make sure they have the same length. After that, they are inputted as features. Different labels are assigned to the data sets corresponding to different gesture intentions. In our case, gesture draw 1 are labeled as 0, gesture draw 2 are labeled as 1 and gesture draw 3 are labeled as 2. Fifty percent of the total data is used for testing, while the other fifty percent is used for training the model. And the model we use is LSTM (Long Short-term Memory) model.

LSTM model is a deep learning model suitable for classifying sequential data. In this study, we implemented a dual-layer Long Short-Term Memory (LSTM) network to handle time series prediction tasks. The model consists of an input layer that takes sequences of shape $(\text{length\_of\_data}, 6)$, where length_of_data is the sequence length and 6 is the number of features (trajectory and orientation in x, y and z plane). The first LSTM layer has 100 units and returns sequences to maintain temporal dependencies, followed by a dropout layer with a dropout rate of 0.2 to reduce overfitting. The second LSTM layer also has 100 units but does not return sequences, preparing the model to finalize output processing. This is

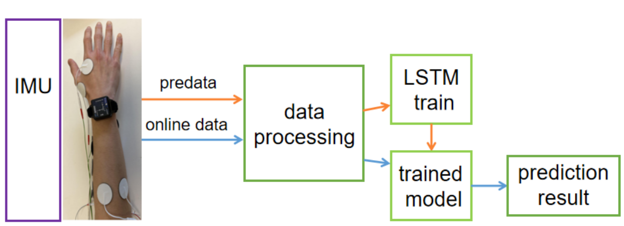

Fig. 2. General process of collecting IMU data to train LSTM model and predict the intention result.

```
model = Sequential()
model.add(LSTM(100, input_shape=(length_of_data, 6),
               return_sequences=True,kernel_regularizer=l2(0.001)))
model.add(Dropout(0.2))
model.add(LSTM(100, return_sequences=False,
               kernel_regularizer=l2(0.001)))
model.add(Dropout(0.2))
model.add(Dense(y.shape[1], activation='softmax',
                kernel_regularizer=l2(0.001)))
```

Fig. 3. LSTM layers used in final online LSTM model traning.

followed by another dropout layer with the same dropout rate. The output layer is a densely connected layer with a softmax activation function, used to output the probability distribution over the target classes. The model uses L2 regularization with a coefficient of 0.001 in the output layer to prevent coefficient inflation. It is compiled with the Adam optimizer and categorical crossentropy loss, and it measures accuracy as a performance metric. The model is trained with a batch size of 10 for 100 epochs, using 50% of the data for testing to validate the model. Upon training completion, the model achieved a test accuracy of 97%. Finally, the model is saved in HDF5 format for subsequent use. This setup highlights the model's robustness in handling overfitting through dropout and L2 regularization, ensuring generalizability on unseen data.

And finally when we came to online gesture recognition using IMU, we add more keinel_regularized: l2(0.001) in our layer as we want to prevent our model from overfitting. The action could greatly increse our accuracy when it comes to online recognition. Our final LSTM model is show in Fig.3 using Keras.

V. EXPERIMENT

A. System Hardware

In this experiment, the most important hardware used is IM900, and then introduce the relevant information of this hardware in detail.IM900 is a state-of-the-art sensing device with robust capabilities, designed to interface seamlessly with personal PCs via BLE 5.0 Bluetooth technology. The IM900 boasts an impressive 8dbm transmission power and an extended communication range of up to 60 meters, ensuring reliable data transfer even over substantial distances. The device's reporting frequency is highly versatile, ranging from 0.5 Hz to 250 Hz; for the purpose of our experiment, we have selected a 30 Hz reporting frequency to optimize the balance between temporal resolution and data volume.

As shown in the picture below, the direction of rotation is defined by the right hand rule, that is, the right thumb points to the axis, and the direction of the four fingers bending is the direction of rotation around the axis.

There are two ways to connect the IM900, one is through the serial port connection, and the other is through Bluetooth connection.When using the serial port connection method, it should be noted that if the Vcc is connected to the serial port tool 5V, when the battery power is not enough, the sensor internal battery will be charged, there will be a charging current within 200ma, the general market serial port tool can not output such a large current, there will be a serial port tool heat or can not be connected to the phenomenon. In addition, when the battery is rapidly charged, the accuracy of the sensor data will decrease, and the accuracy will return to normal after stopping the charging, so the Vcc test is generally not connected. If you need to use vcc for a long time, you can also reduce the charging parameters to prevent the accuracy impact caused by the jitter of charging current.If you are using Bluetooth connection, we only need to use the windows10 laptop with Bluetooth, the desktop can be coupled with a USB Bluetooth adapter, and directly double-click to run the matching IMU upper computer software, you can start to use. Because Bluetooth connection is more convenient and fast, we use Bluetooth connection in the actual experiment.

Our research involved the collection of comprehensive motion data sets using the IM900. This included tri-axial acceleration data, with a full-scale range of ±16g, allowing for the capture of both subtle and dynamic movements. Additionally, tri-axial angular velocity measurements were recorded, with a capacity of ±2000°/s, to accurately document the rotational aspects of limb motion. The inclusion of a tri-axial magnetometer, primarily utilized for automatic calibration, further enriches the data, ensuring high fidelity in the representation of the arm's orientation in space.

The IM900 was strategically positioned at the proximal end of the forearm to ensure unobstructed data collection. This placement was critical in guaranteeing that the device's readings were solely reflective of the arm's motion, without interference from the hand's movements.

B. Calibration

When experimenting with IMUs, calibration is required. Module calibration includes accelerometer calibration, which is not necessary, and magnetic field calibration, which must be calibrated in the use environment if the magnetic field is fused, and Z-axis and coordinate system zero. The accelerometer calibration of the module has been completed before leaving the factory. We only need to check whether the calibration is successful. The criterion for judging is that no matter how the module is placed, the mode value including gravity acceleration is close to 9.8m/s at rest. If the acceleration is found, it needs to be recalibrated. The im900 supports high precision spherical fitting calibration method, which needs to be collected by six surfaces or more data, which can be collected by six surfaces, and the more accurate the data collection. The algorithm will be used for all the data collected, and the optimal calibration parameters should be matched, and the precision calibration is achieved.

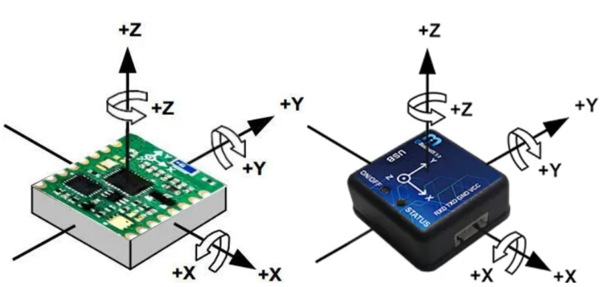

Fig. 4. Sensor axis

C. Data Collection

For data such as numbers 1, 2, and 3, the process is as follows: IM900 is worn on the hand, and then each number is written by hand 20 times, that is, each number has 20 sets of data, and then the model is trained with a 1:1 ratio of training set and test set. For the data of up, down, left and right, we collect them in the following way: Three people collect data, and each person completes each action 20 times while wearing IM900, that is to say, there are a total of 60 sets of data for each action. The reason why we choose multi-person experiment rather than multiple experiments is that different people's movement habits may be different, so as many people as possible to complete, so that the trained model can predict more accurately.

D. Offline and online experiments

For numbers 1, 2 and 3, the intention recognition experiment was offline, that is, IM900 was not connected to receive hand movements in real time for intention recognition, but the collected hand data was used as input, and then the data was classified to determine which number it belonged to. For up, down, left and right gestures, online real-time experiments are used. We hold the IM900 in our hands as we did when we collected the data, and then make one of the actions. This data is transmitted to the computer in real time, and after the prediction of the trained model, the classification result is finally output.

VI. Results and discussion

A. Offline Classification Results

Figure 6 show confusion matrix of gestures 1, 2 and 3 using LSTM model.The labels 0, 1 and 2 represent gestures 1, 2, and 3, respectively.The training set and test set of the LSTM model

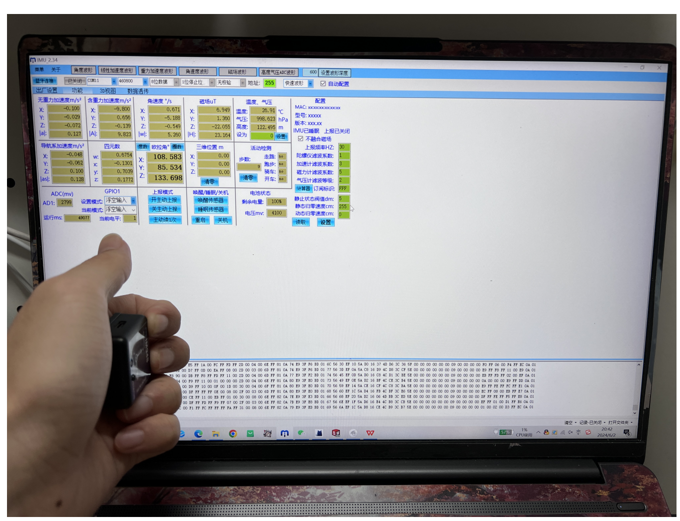

Fig. 5. Data collection

are obtained by collecting the writing gestures 1,2 and 3 of three testers wearing IMU, and the sample size ratio of the test set and training set is 1:1.

From the table, it can be found that 11 of the 12 samples with a true gesture of 1 were correctly predicted by the model LSTM, and one was misjudged as a gesture of 2. The reason for the misjudgment may be that the features of gesture 1 are similar to those of gesture 2, or the subject's hand is unstable when writing gesture 1, and the accuracy of the prediction is still very high.In addition, the 9 samples with a real gesture of 2 and the 6 samples with a real gesture of 3 were all correctly judged by the model LSTM. In general, it can be found that the classification effect of model LSTM is very good through confusion matrix.

B. Online Classification Results

As discussed in the data section, we used gestures such as draw 1, 2, and 3 for offline classification. For online classification, following the advice of Professor Song, we selected gestures: up, down, left, and right. These gestures were collected by moving hands in the corresponding direction while wearing the IMU on the hand. We gathered 20 samples from 3 individuals for each direction, totaling 60 samples per direction and 300 samples overall.

To reduce overfitting, we adjusted our layer structure as shown in Fig.3. For the final online recognition, we used the filtered data provided by the IMU manufacturer, which significantly improved both our offline and online recognition accuracy. As shown in Fig.7, all 47 samples used for testing were predicted correctly, which is a significant proof of our improvement. The training and validation accuracies, as well as losses, are shown in Fig.8 and Fig.9. According to the accuracy figure, after epoch 7, the accuracy stabilized at around 97 percent. Furthermore, there is no significant difference between training and validation accuracy, proving our reduction in overfitting. Finally, in our online recognition,

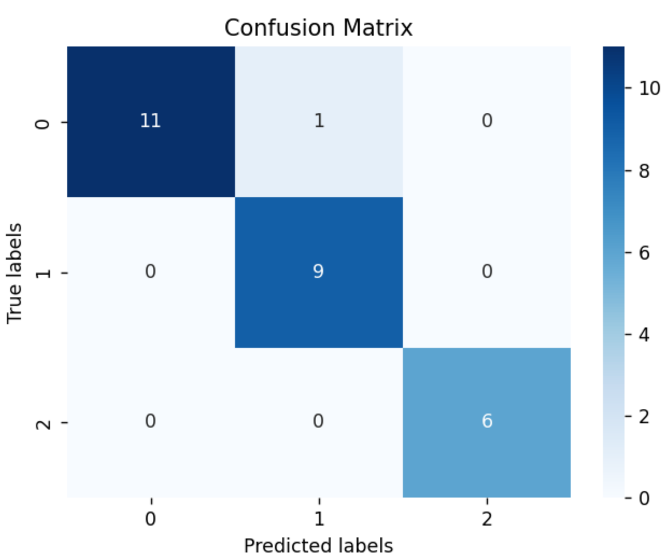

Fig. 6. Offline confusion matrix of gestures 1, 2 and 3 using LSTM model.

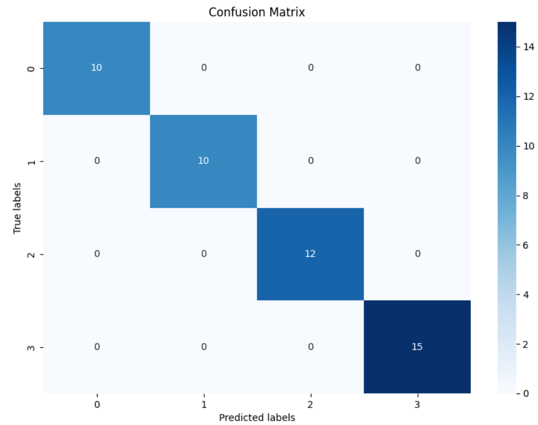

Fig. 7. Online confusion matrix of the online training data. Up is labeled as 0, down is labeled as 1, left is labeled as 2 and right is labeled as 3. And the test size is 20 percent. Training epoch is 30 while batch size is 10.

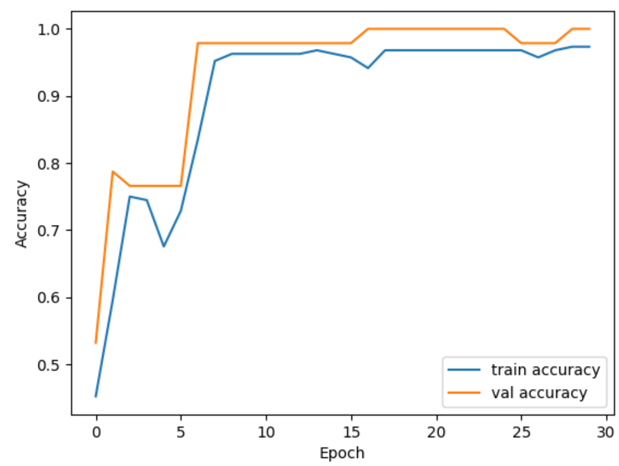

Fig. 8. Train accuracy and validation accuracy changes as epoch increases. eventually came to a stablized probability of about 97 percent.

we were able to clearly distinguish our gestures, achieving an accuracy of approximately 85 percent.

C. Limitation

Although our offline and online accuracy is very high, the gestures we predict are simple and distinct. Considering we only used data collected under 30Hz conditions, this is still a very powerful result. However, more complex gestures may require higher sampling rates and further optimization.

VII. Conclusion

Assistive devices to help people with disabilities to perform grasping manoeuvres are an effective way to help people with disabilities to increase their self-care ability in daily life.Therefore, we constructed models for recognising the arm

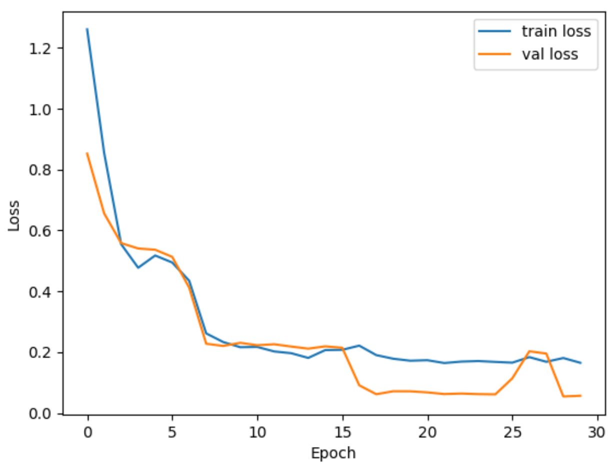

Fig. 9. Train loss and validation loss changes as epoch increases. eventually came to a stablized value of about 0.15

movements of disabled people and controlling the assistive devices by using the arm movements as commands.

We chose a single object for training data and multiple objects for testing the model. Although the amount of training data for our model is small, the accuracy for the test data is 96.3%, which shows that the model out by this method has excellent generality.

With our model, we are able to accurately recognise arm movements not only offline but onine. In the future, we hope to combine this model with robotic arm grasping to send commands through arm movements to control the robotic arm to grasp the corresponding objects, in order to achieve assisting people with disabilities to perform grasping movements completely.

References

[1] S. C. Colachis IV, M. A. Bockbrader, M. Zhang, D. A. Friedenberg, N. V. Annetta, M. A. Schwemmer, N. D. Skomrock, W. J. Mysiw, A. R. Rezai, H. S. Bresler *et al.*, "Dexterous control of seven functional hand movements using cortically-controlled transcutaneous muscle stimulation in a person with tetraplegia," *Frontiers in Neuroscience*, vol. 12, p. 208, 2018.

[2] D. A. Friedenberg, M. A. Schwemmer, A. J. Landgraf, N. V. Annetta, M. A. Bockbrader, C. E. Bouton, M. Zhang, A. R. Rezai, W. J. Mysiw, H. S. Bresler *et al.*, "Neuroprosthetic-enabled control of graded arm muscle contraction in a paralyzed human," *Scientific reports*, vol. 7, no. 1, p. 8386, 2017.

[3] Y. Wang, B. Metcalfe, Y. Zhao, and D. Zhang, "An assistive system for upper limb motion combining functional electrical stimulation and robotic exoskeleton," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 2, pp. 260–268, 2020.

[4] N. Bhagat, K. King, R. Ramdeo, A. Stein, and C. Bouton, "Determining grasp selection from arm trajectories via deep learning to enable functional hand movement in tetraplegia," *Bioelectronic Medicine*, vol. 6, pp. 1–8, 2020.

[5] Y. Zheng and X. Hu, "Elicited finger and wrist extension through transcutaneous radial nerve stimulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 9, pp. 1875–1882, 2019.

[6] D. Yang, Q. Huang, Z. Jiang, and L. Jiang, "Design of multi-channel electrical stimulator integrated with online impedance measurement," *Journal of Medical and Biological Engineering*, vol. 40, pp. 943–950, 2020.

[7] R. A. El-laithy, J. Huang, and M. Yeh, "Study on the use of microsoft kinect for robotics applications," pp. 1280–1288, 2012.

[8] J. Jantz, A. Molnar, and R. Alcaide, "A brain-computer interface for extended reality interfaces," pp. 1–2, 2017.

[9] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, and T. Dutoit, "Performance of the emotiv epoc headset for p300-based applications," *Biomedical engineering online*, vol. 12, pp. 1–15, 2013.

[10] Y. Z. D. Z. Yansong Wang, Benjamin Metcalfe, *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 2, pp. 260–268, 2020.

[11] I. Lopatovska, K. Rink, I. Knight, K. Raines, K. Cosenza, H. Williams, P. Sorsche, D. Hirsch, Q. Li, and A. Martinez, "Talk to me: Exploring user interactions with the amazon alexa," *Journal of Librarianship and Information Science*, vol. 51, no. 4, pp. 984–997, 2019.

[12] A. S. Tulshan and S. N. Dhage, "Survey on virtual assistant: Google assistant, siri, cortana, alexa," pp. 190–201, 2019.