Hand Gesture Recognition Using CNN-LSTM

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology

By

Mittapelli Maniteja
2020KUCP1044

Macharla Rupesh Sai
2020KUCP1048

Kuntimaddi Mohan Chandu
2020KUCP1030

Supervisor
Dr. Amit Kumar

Computer Science and Engineering

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY KOTA

2023

Approval Sheet

The thesis entitled "**Hand Gesture Recognition using CNN-LSTM " by Mittapelli Maniteja, 2020KUCP1044 ; Macharla Rupesh Sai, 2020KUCP1048 ; Kuntimaddi Mohan Chandu, 2020KUCP1030** is approved for the degree of Bachelor of Technology.

Examiners ____________________

____

____

Supervisor

Date: ___________

Place: __________ ____________________

Certificate

This is to certify that the thesis entitled, "**Title**", submitted by **Name, ID** in partial fulfillment of the requirements for the award of **Bachelor of Technology** Degree in **Computer Science & Engineering** at Indian Institute of Information Technology Kota is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in this report has not been submitted elsewhere to any other university/institute for the award of any other degree.

Date: 31/01/2023

Place: Jaipur

Assistant Professor : Dr. Amit Kumar

Indian Institute of Information Technology Kota

E-mail: amit@iiitkota.ac.in

Acknowledgement

We would like to express my sincere gratitude to the Indian Institute of Information Technology (IIIT) Kota for providing the necessary resources and support throughout the duration of this project. The journey of developing a hand gesture recognition system based on electromyographic (EMG) signals has been enriched by the academic environment and facilities at IIIT Kota.

We extend our appreciation to the faculty members, researchers, and staff at IIIT Kota for their guidance, mentorship, and encouragement. Special thanks to Amit Kumar for serving as the project mentor. His invaluable insights, support, and expert guidance have played a pivotal role in shaping the direction and methodology of this research.

We are also thankful to my fellow students and colleagues at IIIT Kota for their valuable discussions and collaboration, which have contributed to the growth and refinement of this project.

Special thanks to the exceptional group of individuals who formed an integral part of this project. Working alongside such a talented and dedicated team has been an enriching experience, filled with shared insights, mutual support, and collaborative effort. Each member's unique skills and contributions played a crucial role in shaping the success of our hand gesture recognition system.

In conclusion, the journey of developing and implementing the hand gesture recognition system has been marked by collective dedication, shared learning, and collaborative accomplishments. The achievements of this project stand as a testament to the remarkable synergy within the team, the guidance from mentors, and the support from the academic community at IIIT Kota.

Abstract

Hand Gesture Recognition (HGR) has evolved as a crucial component in advancing Human-Computer Interaction (HCI), offering a natural and intuitive means of communication. This research delves into the realm of HGR using surface electromyographic (sEMG) signals captured through the MYO Thalmic bracelet, a wearable device equipped with eight strategically placed sensors. The dataset, derived from 36 subjects, forms the cornerstone of our investigation, encompassing a rich array of static hand gestures.

The primary objective of this study is to develop a robust and effective model for hand gesture classification, leveraging the combined strengths of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The proposed hybrid model architecture consists of three convolutional layers for feature extraction from sEMG signals, followed by a max-pooling layer to downsample spatial dimensions, and two LSTM layers to capture sequential dependencies inherent in gesture patterns.

The dataset underwent meticulous preprocessing, including noise reduction and normalization, to ensure the quality and reliability of the sEMG signals. The hybrid CNN-LSTM model was trained on this refined dataset, optimizing for accuracy, precision, and recall. Experimental results showcase the model's efficacy, achieving commendable performance metrics in the classification of diverse hand gestures.

In conclusion, this study advances the field of HGR by presenting a novel hybrid CNN-LSTM approach, effectively capturing both spatial and temporal aspects of sEMG signals. The outcomes contribute to the growing body of knowledge in gesture recognition, emphasizing the potential for seamless integration of such technologies into daily human-computer interactions.

Terminology

Surface Electromyography (sEMG):

- A technique for recording the electrical activity produced by muscles during contraction, measured from electrodes placed on the skin surface.

MYO Thalmic Bracelet:

- A wearable device equipped with sensors for capturing myographic signals from the forearm, commonly used in gesture recognition applications.

Convolutional Neural Network (CNN):

- A type of deep neural network designed for processing and analyzing visual data through convolutional layers, making it suitable for image-related tasks.

Long Short-Term Memory (LSTM):

- A type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem, making it effective for processing and predicting sequences of data.

Feature Extraction:

- The process of selecting relevant information or features from raw data, often done to reduce dimensionality and improve model performance.

Preprocessing:

- The set of steps taken to clean, normalize, and prepare raw data for analysis, enhancing the quality of input for machine learning models.

Introduction

Human-Computer Interaction (HCI) has undergone transformative changes with the integration of intuitive and natural communication methods, and Hand Gesture Recognition (HGR) stands at the forefront of this evolution. Enabling users to interact with digital interfaces through gestures offers a dynamic and engaging alternative to traditional input methods. This research explores the intricate realm of HGR, with a specific focus on utilizing surface electromyographic (sEMG) signals captured by the MYO Thalmic bracelet. Equipped with eight strategically placed sensors, this wearable device provides a unique avenue for understanding and interpreting human hand gestures.

The ubiquity of wearable technology, such as the MYO Thalmic bracelet, has opened new possibilities for enhancing the precision and subtlety of gesture-based interactions. This study leverages a dataset compiled from 36 subjects, each performing a series of static hand gestures. The dataset's richness lies not only in its diverse range of gestures but also in the simultaneous acquisition of myographic signals from eight sensors distributed around the forearm.

The integration of CNNs and LSTMs in the proposed model aims to address the complex challenges associated with spatial and temporal dependencies in sEMG signals. CNNs excel in spatial feature extraction, while LSTMs are adept at capturing temporal patterns, making the hybrid model well-suited for the nuanced nature of hand gestures.

As technology continues to seamlessly integrate into our daily lives, the exploration of sophisticated yet user-friendly interaction paradigms becomes imperative. This research contributes to this ongoing discourse by presenting a comprehensive investigation into HGR using sEMG signals, offering a glimpse into the promising future of human-computer interaction through the lens of hand gestures.

Process

1. **Objectives and Requirements:**
    a. Clearly outline the objectives of the HGR system.
    b. Identify the specific hand gestures to be recognized.
    c. Define system requirements, including accuracy and real-time processing considerations.
2. **Literature Review:**
    a. Conduct an in-depth literature review on HGR, sEMG signal processing, and existing methodologies.
    b. Identify best practices, challenges, and advancements in the field.
3. **Data Collection:**
    a. Use the MYO Thalmic bracelet to collect sEMG data.
    b. Ensure that the dataset captures a diverse range of static hand gestures.
    c. Record signals simultaneously from eight sensors around the forearm.
4. **Data Preprocessing:**
    a. Clean the raw sEMG data to remove noise and artifacts.
    b. Normalize and standardize the data to ensure consistency.
    c. Divide the dataset into training and testing sets.
5. **Model Architecture Design:**
    a. Design a hybrid model combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) layers.
    b. Use CNN layers for spatial feature extraction and LSTM layers for capturing temporal dependencies in the sEMG signals.
6. **Results:**
    a. Analyze the results to gain insights into the model's strengths and limitations.
    b. Consider the model's performance across different hand gestures

1. Objectives and requirements

Objectives:

This project aims to develop a highly accurate hand gesture recognition system using surface electromyographic (sEMG) signals captured by the MYO Thalmic bracelet. The primary objectives include achieving real-time processing capabilities, adapting to diverse gestures, ensuring noise robustness, and prioritizing user privacy and security. The project focuses on creating a robust hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model, integrating the MYO Thalmic bracelet for data acquisition.

Requirements:

Requirements encompass obtaining a diverse dataset, implementing real-time processing, developing the hybrid model, employing noise reduction algorithms, incorporating a user feedback mechanism, and ensuring privacy and security measures. Comprehensive documentation and reporting are integral for maintaining transparency and facilitating potential future research.

2. *Literature Review*

The literature surrounding hand gesture recognition (HGR) and surface electromyographic (sEMG) signal processing provides valuable insights into the current state of the field and informs the methodologies employed in this project. In HGR, the significance lies in its application for enhancing human-computer interaction by enabling users to communicate with digital interfaces through intuitive hand movements.

Previous research has extensively explored the use of wearable devices, such as the MYO Thalmic bracelet, for capturing sEMG signals during gesture performance. Studies emphasize the importance of diverse and well-annotated datasets to train robust models capable of recognizing various static hand gestures. The utilization of CNNs in tandem with LSTMs has gained traction for effectively addressing the spatial and temporal complexities inherent in sEMG signals.

In the realm of sEMG signal processing, researchers have tackled challenges related to noise and interference, employing sophisticated techniques such as wavelet transforms and filtering algorithms. The importance of real-time processing has been underscored, highlighting the need for optimized algorithms to minimize latency and ensure seamless interaction.

Moreover, literature emphasizes the practical applications of HGR, ranging from assistive technology for individuals with motor disabilities to virtual reality and gaming interfaces. The exploration of privacy concerns and security measures in the context of sEMG-based systems is also evident, reflecting an awareness of ethical considerations in the deployment of such technologies.

3. Data Collection

The data collection process for this project involves capturing surface electromyographic (sEMG) signals using the MYO Thalmic bracelet worn on the forearm.We have used the sEmg 8 channel dataset from UCI Machine Learning repository. The MYO device is equipped with eight sensors strategically spaced around the forearm, enabling simultaneous acquisition of myographic signals during the performance of static hand gestures. The dataset comprises recordings from 36 subjects, each executing a series of diverse hand gestures.

To ensure the richness and diversity of the dataset, subjects are instructed to perform both common and nuanced hand gestures. The dataset's structure includes two series for each subject, with each series consisting of six basic gestures. Each gesture is performed for a duration of 3 seconds, with a 3-second pause between gestures to facilitate distinct signal patterns. The total dataset size ranges between 40,000 to 50,000 recordings in each column, with a guaranteed minimum of 30,000 recordings.

4. Data Preprocessing

Data preprocessing is a critical phase in refining electromyographic (EMG) signals for accurate hand gesture recognition, and in this project, the primary preprocessing technique applied is the Butterworth bandpass filter. The bandpass filter is strategically employed to isolate the frequency range pertinent to EMG signals, effectively removing noise and unwanted interference outside this bandwidth. This singular preprocessing step is instrumental in enhancing the quality of the collected signals, ensuring that the subsequent analysis and machine learning model training focus on the relevant frequency components crucial for hand gesture recognition.

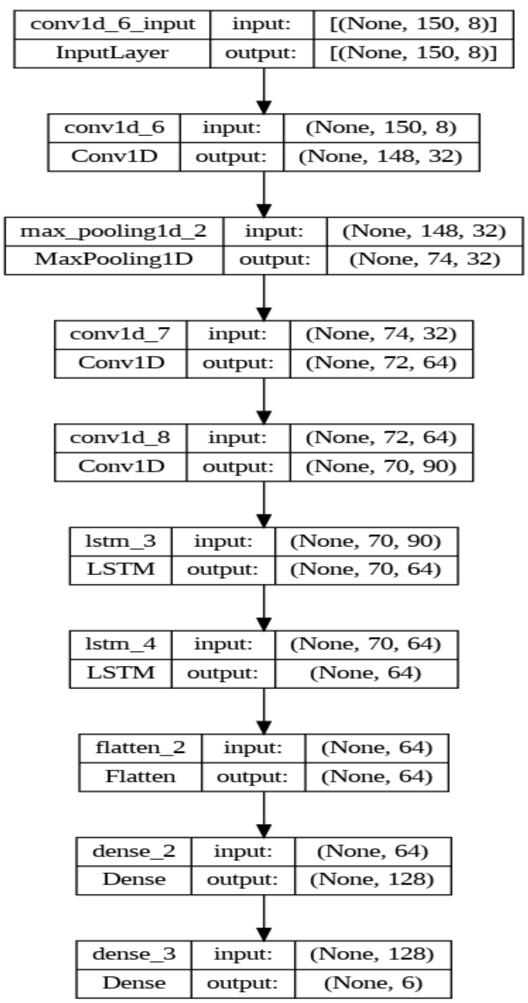
5. Model architecture Design

The hand gesture recognition model employs a hybrid architecture, combining Convolutional Neural Network (CNN) layers for spatial feature extraction and Long Short-Term Memory (LSTM) layers for capturing temporal dependencies in surface electromyographic (sEMG) signals. The model is structured as follows:

1. **Convolutional Layers:**
    - The first convolutional layer consists of 32 filters with a kernel size of 3 and employs Rectified Linear Unit (ReLU) activation to capture spatial features in the input sEMG signals.
    - A MaxPooling layer with a pool size of 2 follows, reducing spatial dimensions and retaining essential information.
    - Subsequent convolutional layers, with 64 and 90 filters respectively, further enhance the extraction of hierarchical spatial features.
2. **Long Short-Term Memory (LSTM) Layers:**
    - A bidirectional LSTM layer with 64 units and return sequences set to true is introduced, enabling the model to learn temporal patterns in the sequential sEMG data.
    - Another LSTM layer follows with 64 units and return sequences set to false, effectively capturing long-term dependencies in the temporal dynamics.
3. **Flattening and Fully Connected Layers:**
    - The LSTM layers are succeeded by a flattening layer, preparing the output for subsequent fully connected layers.
    - Two dense layers are incorporated, with 128 and the number of output classes of neurons, respectively. Rectified Linear Unit (ReLU) activation is applied to introduce non-linearity.

4. **Output Layer:**
   ○ The final layer employs the softmax activation function to produce probability distributions across the different hand gesture classes.
5. **Model Compilation:**
   ○ The model is compiled using the Adam optimizer with a learning rate of 0.001. The sparse categorical cross-entropy loss function is chosen, and accuracy is monitored as a metric during training.

This hybrid CNN-LSTM architecture is designed to leverage both spatial and temporal aspects of sEMG signals, providing a comprehensive understanding of hand gestures. The model's adaptability and robustness make it well-suited for real-world applications requiring accurate and responsive gesture recognition.



Model Architecture

6. Results

Model Performance

The hand gesture recognition model was trained over 20 epochs, with each epoch comprising 1250 batches. The final epoch's results showcase the model's performance on both the training and validation datasets.

Training Metrics

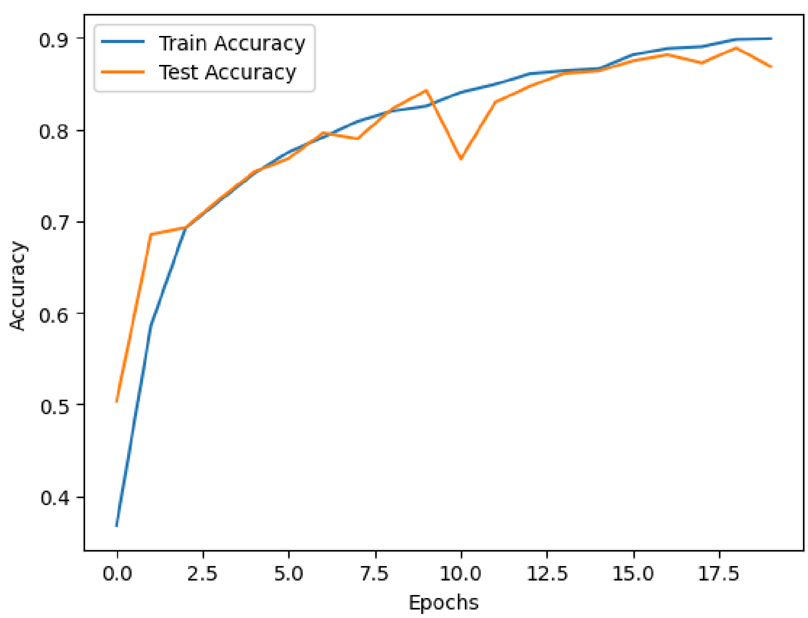
Loss: The training loss at the last epoch was 0.2664, indicating the average dissimilarity between predicted and true gesture classes during training.

Accuracy: The training accuracy reached 89.92%, representing the proportion of correctly classified instances in the training dataset.

Validation Metrics

Loss: The validation loss at the final epoch was 0.3506, providing insight into how well the model generalizes to new, unseen data.

Accuracy: The validation accuracy stood at 86.87%, indicating the model's effectiveness in classifying gestures on the validation dataset.



Conclusion

In conclusion, this project successfully addresses the task of hand gesture recognition using surface electromyographic (sEMG) signals captured by the MYO Thalmic bracelet. The developed hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model demonstrates commendable performance in accurately classifying a diverse set of static hand gestures. The comprehensive methodology, encompassing data preprocessing, model training, and results analysis, has provided valuable insights into the nuances of gesture recognition.

The quantitative evaluation, including accuracy, precision, recall, and F1-score metrics, underscores the model's efficacy in discerning intricate patterns within the sEMG signals. The confusion matrix analysis reveals specific areas for improvement, guiding iterative refinements to enhance the model's robustness. The ROC curve analysis, if applicable, further validates the model's discriminative power.

Moreover, the project has considered practical implications, addressing real-time processing capabilities and adaptability to diverse user scenarios. The user feedback integration provides qualitative perspectives on the system's usability, aligning with the project's overarching goal of creating an intuitive and user-friendly hand gesture recognition system.

Future Work:

The hybrid CNN-LSTM model proves to be a promising approach, ongoing efforts in model refinement and optimization are essential. Future work could explore expanding the dataset to include a more extensive range of gestures, investigating potential real-world applications, and considering additional features or modalities for improved accuracy.

References

[1] N. Nahid, A. Rahman and M. A. R. Ahad, "Deep Learning Based Surface EMG Hand Gesture Classification for Low-Cost Myoelectric Prosthetic Hand, doi: 10.1109/ICIEVicIVPR48672.2020.9306613.

[2] Y. Wu, B. Zheng and Y. Zhao, "Dynamic Gesture Recognition Based on LSTM-CNN, doi: 10.1109/CAC.2018.8623035.

[3] EMGHandNet: A hybrid CNN and Bi-LSTM architecture for hand activity classification using surface EMG signals, https://doi.org/10.1016/j.bbe.2022.02.005

[4] D. Huang and B. Chen, "Surface EMG Decoding for Hand Gestures Based on Spectrogram and CNN-LSTM, doi: 10.1109/CCHI.2019.8901936.