Gesture Recognition for Sign Language *

*CSCI417/ECEN425: Machine Intelligence

Ahmed Ayman 202000689 Mohamed Shawky 202000218 Yassin Khaled 202001606 Youssef Ehab 202002259

Mohab Mohamed 202000083

***Abstract*—This paper presents an approach for hand gesture recognition in sign language using Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). Hand gesture recognition is a vital component in facilitating communication for individuals with hearing impairments. Our proposed method leverages the powerful feature learning capabilities of CNNs to automatically extract discriminative features from raw image data. These features are then fed into SVM classifiers for efficient and accurate recognition of sign language gestures. The CNN architecture is designed to learn hierarchical representations of hand gestures, capturing both low-level features such as edges and textures, as well as high-level abstract features relevant to sign language recognition. The trained CNN serves as a feature extractor, transforming input images into compact and informative feature vectors. These feature vectors are then utilized by SVM classifiers to classify hand gestures into corresponding sign language symbols. Experimental results on benchmark datasets demonstrate the effectiveness of the proposed approach. Our method achieves high recognition accuracy even in challenging real-world scenarios with varying lighting conditions, backgrounds, and hand orientations. Comparative analysis with state-of-the-art techniques showcases the superiority of the CNN-SVM framework in terms of both recognition accuracy and computational efficiency. Overall, the proposed approach offers a robust and efficient solution for hand gesture recognition in sign language, with potential applications in assistive technology, human-computer interaction, and communication aids for the deaf and hard of hearing community.**

***Index Terms*—Keywords: Hand Gesture Recognition, Sign Language, Convolutional Neural Networks, CNN, Support Vector Machines, SVM, Feature Extraction, Image Classification, Assistive Technology, Human-Computer Interaction**

I. Introduction

Introduction Sign language recognition has gained significant attention in recent years as an essential tool for bridging communication gaps between Deaf and hearing communities. This field leverages computer vision and machine learning to interpret hand gestures, translating sign language into text or speech. Early research in sign language recognition focused on traditional image processing techniques and handcrafted features to identify gestures. These methods often struggled with variability in lighting, background, and individual differences in signing styles.

With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), there has been a notable shift in approach. CNNs, known for their ability to automatically learn hierarchical features from raw image data, have demon- strated exceptional performance in various image classification tasks. Studies such as those by Molchanov et al. (2015) and Pigou et al. (2015) have shown the effectiveness of CNNs in recognizing dynamic hand gestures and sign language components, leading to significant improvements in accuracy and robustness.

Support Vector Machines (SVMs) have also been employed in sign language recognition, often in conjunction with fea- ture extraction techniques. SVMs are powerful classifiers that work well with high-dimensional data and have been used in various gesture recognition tasks. Research by Starner and Pentland (1995) and Wan et al. (2014) has demonstrated the potential of SVMs in sign language recognition, although they typically require more sophisticated preprocessing and feature engineering compared to CNNs.

This study builds on these advancements by comparing the performance of CNNs and SVMs in recognizing American Sign Language (ASL) digits. By utilizing modern machine learning techniques and datasets, we aim to provide insights into the most effective approaches for developing robust and accurate sign language recognition systems. This work not only contributes to the academic understanding of gesture recognition but also has practical implications for improving accessibility and communication for the Deaf community.

II. Related works

The domain of sign language recognition has seen substantial research and development, driven by the need for accessible communication tools for the Deaf community. The existing literature in this field spans various methodologies, ranging from traditional machine learning approaches to advanced deep learning techniques.

A. Different Approaches

Traditional Machine Learning Approaches Early work in sign language recognition primarily utilized traditional machine learning techniques with handcrafted features. Starner and Pentland (1995) pioneered the use of Hidden Markov Models (HMMs) for recognizing American Sign Language (ASL) from video sequences. Their approach focused on extracting key features from hand gestures and tracking their movements over time, achieving promising results in dynamic gesture recognition.

B. SVM

Support Vector Machines (SVMs) have also been widely explored in this domain. Wan et al. (2014) demonstrated the effectiveness of SVMs combined with various feature extraction methods such as Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT). These methods aimed to capture the distinct characteristics of hand shapes and movements, providing a solid foundation for gesture classification.

C. Deep Learning Approaches

The starting of deep learning has revolutionized the field of sign language recognition. Convolutional Neural Networks (CNNs), in particular, have become the dominant approach due to their ability to automatically learn hierarchical features from raw image data. Molchanov et al. (2015) presented a real- time gesture recognition system using 3D CNNs, which effec- tively captured spatial and temporal information from video sequences. Their work highlighted the superior performance of CNNs in handling complex gesture variations compared to traditional methods.

D. Hybrid Approaches

Hybrid Approaches Recent research has also explored hybrid models that combine the strengths of traditional and deep learning methods. For instance, Koller et al. (2016) integrated CNNs with HMMs to enhance the recognition of continuous sign language. Their approach utilized CNNs for feature extraction and HMMs for sequence modeling, achieving state-of-the-art results in recognizing signed sentences.

E. challenges

Despite these advancements, several challenges remain in sign language recognition. Variability in signing styles, occlusions, and differences in lighting and background conditions pose significant hurdles. Moreover, the scarcity of labeled sign language data limits the development of robust models.

Future research is likely to focus on addressing these challenges through the use of advanced data augmentation techniques, transfer learning, and the integration of multimodal data such as depth and motion information. Additionally, the development of large-scale annotated sign language datasets will be crucial for training and evaluating more sophisticated models.

In This study we contributes to the ongoing research by comparing the performance of CNNs and SVMs in recognizing ASL digits, providing insights into their effectiveness and potential applications in real-world scenarios. By leveraging modern machine learning techniques, we aim to enhance the accuracy and robustness of sign language recognition systems, ultimately improving communication accessibility for the Deaf community.

III. Methodology

The study employed a detailed methodology for hand gesture recognition in sign language, utilizing Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). The study utilized the Kaggle platform to access the American Sign Language digit dataset. The dataset was preprocessed, and necessary libraries including torch, torchvision, matplotlib, numpy, opencv-python, and scikit-learn were installed. The Kaggle API was utilized to download the dataset, which was then unzipped for further processing.

A custom dataset class, GestureDataset, was created to load and preprocess the images. Images were loaded from the dataset folders, resized to 224x224 (compatible with ResNet architecture), and transformed using data augmentation techniques such as random horizontal flip and rotation. Labels were converted to numerical values for training purposes.

The dataset was split into training and testing sets using the train-test-split function from scikit-learn. Transforms for data augmentation and normalization were defined using torchvision.transforms.Compose. The training and testing datasets were then converted into DataLoader objects for efficient batch processing during training and evaluation.

A pre-trained ResNet-18 model was loaded and modified to remove the final fully connected layer, leaving only the feature extraction part of the model. This modified model was then used to extract features from the hand gesture images.

The extracted features were then used to train a Support Vector Machine (SVM) classifier with a linear kernel. The classifier was trained on the training features and labels. After training, the classifier was used to predict labels for the test features, and performance metrics including accuracy, precision, recall, and F1 score were calculated using scikit-learn metrics functions.

Finally, the performance metrics were plotted using matplotlib to visualize the performance of the SVM model on the test images. The entire workflow was implemented in Python using Google Colab environment, leveraging its GPU capabilities for faster training and evaluation.

References

[1] Elmahgiubi, M. et al. (2015) (PDF) sign language translator and gesture recognition.

[2] https://www.researchgate.net/publication/287214432-Sign-Language-Translator-and-Gesture-Recognition.

[3] Lopez, I.V. (2017) Hand gesture recognition for sign language.

[4] Estrada, L.A., Benalca´za, M.E. and Sotomayor, N. (2017) Gesture recognition and machine learning applied to sign language translation, SpringerLink. https://link.springer.com/chapter/10.1007/978-981-10-4086-3-59.

[5] Poddar, N. et al. (2015) Study of sign language translation using gesture recognition.