Sign Language Recognition using Computer Vision

Triboi Maria-Emanuela
Faculty of Automatic Control and Computer Science
Technical University Gheorghe Asachi Iasi
Iasi, Romania, 700050
Email: maria-emanuela.triboi2@student.tuiasi.ro

***Abstract*—Sign language plays a vital role in facilitating communication for individuals with hearing impairments. With the advancements in computer vision and deep learning, sign language detection using computer vision techniques has emerged as a promising approach. This research paper aims to explore the application of computer vision techniques in sign language detection and investigates the challenges and advancements in this field. The paper presents an overview of the current state-of-the-art methods and discusses the datasets and evaluation metrics.**

I. Introduction

Sign language is a visual language that utilizes hand gestures, facial expressions, and body movements to convey meaning and information. It serves as the primary mode of communication for individuals with hearing impairments. However, the ability to understand and interpret sign language is not universal, posing a significant barrier to effective communication. Computer vision techniques offer a promising solution by enabling the development of automated sign language detection systems that can translate sign language gestures into text or speech.

II. Sign Language Recognition Methods

Various methods have been proposed for sign language detection using computer vision techniques. These methods can be broadly categorized into two types: traditional machine learning-based methods and deep learning-based methods. Traditional methods often rely on handcrafted features, such as shape, motion, and appearance-based features, combined with machine learning algorithms. In recent years, deep learning approaches, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown remarkable success in sign language recognition tasks.

A. Template Matching

Template matching is a traditional machine learning-based method for sign language detection [1]. It involves comparing the input image with predefined templates of sign language gestures. The similarity between the input image and the templates is measured using various techniques such as correlation or distance metrics. However, template matching has limitations in handling variations in hand shapes, occlusions, and inter-class similarities. It relies heavily on accurate alignment and scaling of templates, making it less robust in real-world scenarios.

B. Convolutional Neural Networks (CNNs)

In recent years, deep learning approaches, particularly Convolutional Neural Networks (CNNs), have shown remarkable success in sign language recognition tasks [2]. CNNs have the ability to automatically learn discriminative features from raw image data, alleviating the need for handcrafted features. The hierarchical structure of CNNs enables them to capture both low-level and high-level visual patterns, making them suitable for sign language detection.

CNNs for sign language recognition typically consist of multiple convolutional layers for feature extraction, pooling layers for spatial downsampling, and fully connected layers for classification. By training CNNs on large-scale sign language datasets, they can learn to differentiate between different sign gestures, considering variations in hand shapes, movement speeds, and facial expressions. Additionally, data augmentation techniques, such as rotation, translation, and flipping, can be applied to enhance the model's generalization capabilities.

C. Challenges in Sign Language Detection

Sign language detection presents several challenges due to the complexity and variability of hand gestures, occlusions, and inter-class similarities. The variability in hand shapes, movement speeds, and facial expressions adds further complexity to the task. Additionally, occlusions caused by clothing, accessories, or other body parts can hinder accurate detection. Furthermore, distinguishing between visually similar sign gestures requires robust feature representation and discrimination capabilities.

III. Experimental Evaluation

In this section, we evaluate the performance of our proposed sign language detection system using computer vision. We conducted extensive experiments on a dataset of ASL alphabet signs to assess the effectiveness of our approach. Additionally, we provide details of the dataset used and the evaluation metrics employed to measure the system's performance.

A. Dataset description

To assess the effectiveness of the approach, I conducted extensive experiments on the ASL Alphabet Dataset [3]. This dataset includes a collection of images depicting the ASL alphabet signs performed by multiple individuals. It covers all 26 letters of the alphabet (A-Z), providing a diverse set of examples for training and evaluation.

B. Experimental Setup

1) Preprocessing: Before feeding the images into the CNN model, I performed preprocessing steps to ensure consistency and enhance the quality of the data. The images were resized to a uniform size of 64x64 pixels, converted to grayscale, and pixel values were normalized in the range [0, 1]. These preprocessing steps help in reducing variations and standardizing the input data.

2) Model Architecture: The sign language detection system is based on a convolutional neural network (CNN) architecture. The model consists of multiple layers, including convolutional layers for feature extraction, pooling layers for spatial downsampling, and fully connected layers for classification. Dropout layers were incorporated to reduce overfitting.

The CNN model architecture was designed to learn discriminative features from the ASL alphabet images and classify them into their corresponding classes.

3) Training and Validation: To train and evaluate the model, I split the dataset into training and validation sets, allocating 80% of the data for training and the remaining 20% for validation. The training set was used for model parameter optimization, while the validation set allowed me to monitor the model's performance and prevent overfitting.

During training, I employed data augmentation techniques such as rotation, translation, and flipping to augment the training set and improve the model's generalization capabilities. This approach helped the model learn robust and invariant representations of the ASL alphabet signs.

C. Performance Evaluation

After training the model, I evaluated its performance on a separate test dataset. The test dataset consisted of real-world ASL alphabet images that were not seen by the model during training and validation.

I measured the performance of the sign language detection system using various evaluation metrics, including accuracy, precision, recall, and F1-score. These metrics provide insights into different aspects of the model's performance, such as overall correctness, precision of individual classes, and ability to retrieve relevant classes.

D. Results

The experimental results of my sign language detection system initially showed limited accuracy, achieving only approximately 11%. While these results indicate the model's current limitations in correctly recognizing ASL alphabet signs from unseen test images, they provide valuable insights into areas that require improvement.

The utilization of the ASL Alphabet Dataset [3] as a training and evaluation resource allowed for an understanding of the model's generalization and robustness. Despite the initial low accuracy, the diverse collection of ASL alphabet signs performed by multiple individuals in the dataset provided valuable training examples for the model to learn discriminative features.

To improve the accuracy of the sign language detection system, further testing and experimentation are required. By analyzing the shortcomings of the initial model, it is possible to identify areas that can be enhanced. These improvements may include refining the data augmentation strategies, exploring alternative network architectures, and fine-tuning hyperparameters through systematic experimentation.

By conducting additional testing and refining the model, it is expected that its accuracy can be significantly improved. The iterative process of testing, analyzing results, and making adjustments will allow for the development of a more reliable and precise sign language detection system.

E. Enhancements

In order to enhance the accuracy of the sign language detection system, several improvements were implemented. Firstly, a more advanced data augmentation strategy was employed during the training phase, including techniques such as scaling, rotation, and random cropping. This augmentation helped to increase the diversity of the training data and improve the model's ability to generalize to different hand gestures and variations. Secondly, a deeper CNN architecture with increased model capacity was utilized, allowing for the extraction of more complex and discriminative features from the sign language images. This deeper architecture included additional convolutional and fully connected layers, enabling the model to capture more intricate spatial patterns and relationships.

In addition to the mentioned improvements, dynamic adjustments to the learning rate were implemented to further enhance the accuracy of the sign language detection system. This approach aimed to optimize the training process based on the results obtained in the previous epoch.

During training, the model's performance on the validation set was monitored after each epoch. By evaluating the validation metrics, such as accuracy or loss, the training process could adapt dynamically.

Additionally, the learning rate, which determines the step size taken during gradient descent optimization, was adjusted dynamically. Higher learning rates allow for larger updates to the model's parameters but may risk overshooting the optimal values. Lower learning rates result in smaller updates but may lead to slower convergence. By monitoring the validation metrics, the learning rate could be adjusted accordingly to ensure efficient training and prevent suboptimal solutions.

The dynamic adjustment of the learning rate based on the results in the previous epoch helped fine-tune the training process and optimize the model's performance. This iterative approach allowed for better convergence and improved accuracy in the sign language detection system.

IV. Conclusion

In conclusion, this research paper explored the application of computer vision techniques in sign language detection. The advancements in deep learning, particularly CNNs, have significantly improved the accuracy and performance of sign language detection systems. While template matching approaches provide a traditional solution, CNNs offer a more robust and automated approach by learning features directly from the data. The availability of comprehensive sign language datasets, such as the ASL Alphabet Dataset, has been instrumental in advancing this field.

Moving forward, further research is needed to address challenges such as occlusions, inter-class similarities, and variability in hand gestures. Additionally, exploring real-time sign language detection and expanding the system's capabilities to recognize more complex sign language sentences are exciting avenues for future investigation. By harnessing the power of computer vision and deep learning, we can continue to enhance communication and accessibility for individuals with hearing impairments.

ACKNOWLEDGMENT

I would like to express my gratitude to the research community in the field of computer vision and sign language detection for their valuable contributions. I am also thankful for the availability of publicly accessible sign language datasets and the efforts of researchers and organizations in promoting inclusivity and accessibility.

REFERENCES

[1] S. Shrenika and M. Madhu Bala, "Sign Language Recognition Using Template Matching Technique," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2020, pp. 1-5, doi: 10.1109/ICCSEA49143.2020.9132899.

[2] T. Li, Y. Yan and W. Du, "Sign Language Recognition Based on Computer Vision," 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2022, pp. 927-931, doi: 10.1109/ICAICA54878.2022.9844497.

[3] https://www.kaggle.com/datasets/debashishsau/aslamerican-sign-language-aplhabet-dataset