Exploring the Impact of Image Brightness on Sign Language Recognition Using Convolutional Neural Network

Jiawei Zhang
School of Information Engineering, Changchun College of Electronic Technology, Changchun, Jilin, 130000, China

Keywords: Sign Language Recognition, Convolutional Neural Networks, Image Brightness.

Abstract: In order to enhance communication for the hard of hearing, sign language recognition technology is intended to understand sign language motions and convert them into text or voice. The primary goal of sign language recognition technology is to give deaf and normal individuals a means of communicating through signals that is both practical and efficient. Research on sign language identification is ongoing due to advancements in computer technology and the growing popularity of intelligence. Wearable input devices and Convolutional Neural Networks (CNN) are two major machine vision-based research methodologies used today. Strap-on input device-based sign language recognition has an advantage over machine vision-based recognition in that it can acquire real-time information on hand shape, finger flexion, and abduction. The research employs machine learning algorithms to analyze how variations in image brightness can affect the performance of CNNs in interpreting sign language gestures. The study adjusts brightness levels to assess how they impact recognition metrics such as accuracy, precision, recall, and F1 score. The findings suggest that variations in brightness have an impact on the models' accuracy of recognition.

1 INTRODUCTION

The importance of sign language for deaf people's daily life and communication cannot be overstated (Kyle, 1998). First of all, sign language is the main form of communication for those who are hard of hearing. It is a visual language conveyed through body language, facial emotions, and gestures (Stokoe, 2005). Sign language is a means of communication and a way for people who are deaf to express their culture and sense of self. According to the survey data, in Hong Kong, China, more than four-fifths (83.1%) of the hearing-impaired persons were aged 60 or above, which indicated that most of the hearing-impaired persons were elders who relied on television and the Internet in particular for their information acquisition needs (Centre, 2011). Furthermore, over 90% of the hard of hearing people said that sign language interpreters were crucial to comprehending the news conference's content. These figures demonstrate how crucial sign language interpreters are to providing information access for the hard of hearing. Interpreting in sign language is not only something you do in everyday discussions; it's also utilized extensively in a lot of other sectors, like law, healthcare, and education. For example, in courtrooms, sign language interpreters ensure that deaf people can understand and participate in the proceedings (Rosen, 2010). In schools, sign language teachers make knowledge easy to understand and memorize through lively and interesting sign language teaching, thus increasing students' interest and learning effect. However, sign language interpreters also face some challenges. For example, some sign language interpreters may be misunderstood by deaf people because there are differences in sign languages in different regions. Thus, it is especially crucial to have a single sign language standard in order to safeguard the legal rights of the deaf population. The United Nations has also recognized Sign language's significance and established the international sign language day to celebrate and uphold the rights of the deaf. Sign language is not only a communication bridge for deaf people, but also an important part of social inclusion and accessibility. In short, Deaf people's lives are greatly impacted by sign language, which not only makes it easier for them to communicate with hearing people, but also providing them with ways to obtain information, receive education and participate in social life. With the development of technology and social progress, the quality of life for the deaf and

Zhang, J.
Exploring the Impact of Image Brightness on Sign Language Recognition Using Convolutional Neural Network.
DOI: 10.5220/0013296000004558

In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 229-233
ISBN: 978-989-758-738-2

dumb will be significantly improved by the progressive improvement of accessible services and sign language interpretation.

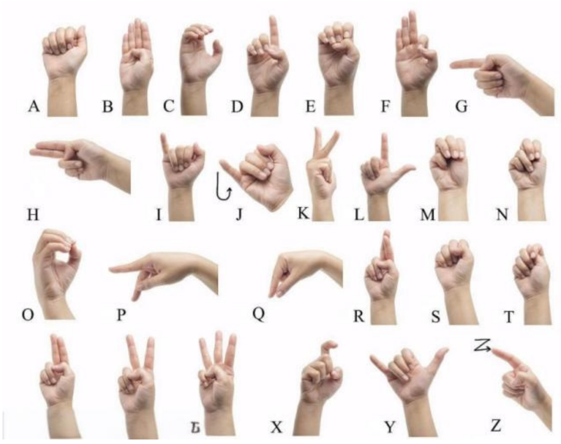

Figure 1: Representative example of American sign language (Bantupalli, 2019).

In order to make the system more intuitive and user-friendly for both the hearing community and those with hearing impairments, this paper will discuss the benefits and drawbacks of different sign language interpreters as well as investigate how brightness affects recognition outcomes from a luminance perspective. The experiments are conducted on American sign language dataset (Bantupalli, 2019) as seen in Figure 1. The effect of luminance on recognition results is examined from the perspective of brightness.

2 METHOD

Convolutional neural networks, or CNNs, are a widely respected family of neural networks in the field of picture identification and classification. well esteemed in the field of classification and image recognition (Alzubaidi, 2021). CNNs use multilayer perceptrons and require only minimal preprocessing to train architectures to perform image recognition and classification tasks, a representative example is demonstrated in Figure 2. With only minimal preprocessing, it is possible to train very efficient architectures for classification tasks. CNNs are designed to resemble the structure of neuronal connections found in the visual brain of an animal (Bhatt, 2021). patterns of neuronal connections in the visual brain of animals. In the realm of image and video recognition, CNNs frequently perform better than other methods. Algorithms for processing images and videos tend to perform better in domains like natural language processing, medical image analysis, and image classification.

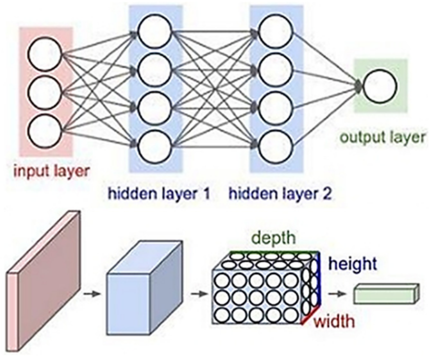

Figure 2: A representative example of CNN (Bantupalli, 2019).

2.1 Key Actions in CNN

There are four key actions in CNN, including convolution, Rectified Linear Unit (ReLU), pooling, and fully connection (Salehi, 2023).

The convolution operation is fundamental for feature extraction. To calculate the dot products between the kernel and the image patches, the input image is covered by a filter, sometimes known as a kernel. Through the collection of local information, features such as edges, textures, and patterns are identified in the image. The resulting feature maps provide a multi-dimensional representation of the input, highlighting different aspects that are crucial for image recognition tasks (Dhruv, 2020).

After the convolution, ReLU is employed to introduce non-linearity into the model. The definition of the ReLU function is

$$f(x)=\max(0, x) \tag{1}$$

, where x is the input feature in this case. It effectively thresholds the input at zero. This implies that positive values in the feature map remain unaltered and that any negative values are set to zero. The network may learn more intricate representations and interactions within the data thanks to the inclusion of non-linearity, which is crucial. ReLU is also computationally effective and aids in addressing the vanishing gradient issue that arises in deep networks with saturated activation functions like tanh or sigmoid.

Subsampling, or pooling, is a technique used to decrease the feature maps' spatial dimensions. By doing this crucial step, the network's sensitivity to the precise placement of features inside the input image will be reduced, improving the model's translational invariance. When pooling techniques like maximum, average, or sum pooling are applied to a portion of the feature map, less significant aspects are removed while the most crucial ones are retained. This reduction in dimensionality also helps to reduce the number of parameters and processing load in the subsequent layers of the network.

A CNN's completely linked layer is essential to its latter phases. In order to carry out tasks like classification or regression, the completely interconnected layers come together and synthesize the collected features after the convolutional and pooling layers have processed the input. Complex decision-making processes are made possible in these layers due to the fact that all of the neurons in the layer above are connected to one another. When it comes to categorization jobs' output layer, activation functions such as SoftMax are frequently employed because they transform the raw output scores of the network into probabilities that add up to one, signifying the chance of each class.

Backpropagation, a crucial technique for training neural networks, is also used at this layer to modify the weights and biases in response to variations in the expected and actual outputs.

In summary, the convolution operation in CNNs captures local patterns within the input, ReLU introduces non-linearity for complex function approximation, pooling reduces the spatial dimensions and computational requirements, and fully linked layers combine the qualities needed for advanced jobs like categorization. Together, these operations form the backbone of CNNs, enabling them to effectively process and learn from image data.

2.2 LeNet Architecture

One of the first CNNs to be widely used was LeNet, which opened the door for other studies on CNNs and multilayer perceptrons (LeCun, 1998). Yann LeCun's groundbreaking creation, LeNet5, is the outcome of several fruitful iterations since 1988. LeNet was mainly created for character recognition applications, such postal codes and digits. activities involving character recognition, such zip codes. Since then, every newly suggested neural network design has had its accuracy evaluated using the Mixed National Institute of Standards and Technology (MNIST) dataset, which has been established and used as a standard.

3 EXPERIMENT AND RESULTS

3.1 Dataset

According to Ayush, while it has a different syntax from English, American Sign Language (ASL) is a complete natural language with many of the same linguistic properties as spoken language (Ayush, 2019). Hand and facial motions are used in ASL communication. It is the primary language of many deaf and hard of hearing North Americans, in addition to being spoken by many others who can hear normally. The dataset's format closely resembles that of traditional MNIST. For every letter A through Z, a label (0–25) is represented in each training and test case, acting as a one–to–one mapping (Neither 9=J nor 25=Z has ever occurred because of gestural movement). The training data (27,455 examples) and test data (7,172 cases) are almost half the size of typical MNIST but otherwise comparable to standard MNIST. The header rows are designated pixel1, pixel2, ..., pixel784, representing a 28x28 pixel image with grayscale values between 0-255. The raw gesture image data represents several people repeating gestures in different conditions. A remarkable expansion of a small batch of color photographs (1704) that were not cropped around the hand region of interest is where the MNIST data for sign language originates.

This work augments the data by resizing, grayscale scaling, cropping to the hand, and then constructing more than 50 versions to increase the number.

The idea of brightness adjustment is to change the amount of light in a picture. This is achieved by altering the pixel values, which range from 0 (black) to 255 (white). By applying a luminance factor to these values, the image can be made brighter or darker. Increasing the factor lightens the image, while decreasing it darkens it. This process is crucial for enhancing image recognition in machine learning models, as it can significantly impact the performance of algorithms like CNNs.

3.2 Performance Comparison

In this work, as demonstrated in Table 1, by modifying the brightness of the dataset, the author gets the following results.

From this result it could be observed that the accuracy of the model for recognizing images increases as the brightness increases. The model at higher and lower brightness decreases the recognition of the model considerably. Since the method value changes the brightness value without controlling for other factors to remain constant, there will be some error.

Table 1: Performance comparison using different brightness.

Brightness	Accuracy	Precision	Recall	F1
-100	0.83	0.91	0.82	0.85
-50	1.00	1.00	1.00	1.00
0	1.00	1.00	1.00	1.00
+50	0.99	0.99	0.99	0.99
+100	0.78	0.87	0.77	0.80

4 DISCUSSIONS

The results indicate a significant correlation between image brightness and recognition accuracy, with optimal performance observed at a luminance level of 0. This implies that under typical illumination circumstances, the CNN model performs best when it comes to identifying sign language motions. The consistency in results within the -50 to 0 luminance range implies a threshold beyond which minor adjustments in brightness have minimal effects on recognition performance. This result could be explained by the model's flexibility within a specific brightness range, after which the changes become less discernible, and thus, less influential on the model's predictive capabilities.

However, the decrease in accuracy at both extreme ends of the brightness spectrum, specifically at -100 and +100, highlights the model's sensitivity to overly dark or bright images. This could be due to the loss of detail and contrast in the images, which are critical for the CNN to distinguish between different sign language gestures.

The study's shortcomings, such as the comparatively limited dataset utilized for the tests, must also be taken into account A more extensive and varied dataset might be advantageous for future study, it could make the findings more broadly applicable and highlight more complex patterns. A more thorough knowledge of the circumstances that maximize sign language recognition might be obtained by investigating additional variables including contrast, color balance, and picture resolution that may have an impact on recognition accuracy.

The talk concludes by highlighting the significance of luminance in sign language recognition systems and the necessity of more research to improve and optimize CNN models' performance under various lighting scenarios. The larger objective of creating inclusive and accessible technology for the deaf and hard-of-hearing community is furthered by this study.

5 CONCLUSIONS

The role of brightness in CNN-based sign language recognition has been clarified in significant ways by this work. The results of the study indicate that the brightness levels of the images used in the training and testing of the CNN model significantly affect the recognition accuracy of sign language.

The optimal performance was achieved with images at a luminance level of 0, suggesting that standard lighting conditions are most conducive to accurate recognition. The findings emphasize the importance of image preprocessing. In this work, brightness adjustments can improve CNN performance in sign language recognition applications. The outcomes also show how important it is to investigate the implications of additional image processing methods that can raise identification rates.

REFERENCES

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*, 1-74.

Ayush, T. 2019, American Sign Language Dataset URL: https://www.kaggle.com/datasets/ayuraj/asl-dataset. Last Accessed: 2024/09/10

Bantupalli, K., & Xie, Y. 2019. American sign language recognition using machine learning and computer vision.

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... & Ghayvat, H. 2021. CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics*, *10*(20), 2470.

Centre H. P. 2011. Be Aware of Hearing Loss, URL: https: //www.chp.gov.hk/files/pdf/ncd_watch_mar_2011_en g.pdf. Last Accessed: 2024/09/10

Dhruv, P., & Naskar, S. 2020. Image classification using c onvolutional neural network (CNN) and recurrent neur al network (RNN): A review. *Machine learning and in formation processing: proceedings of ICMLIP 2019*, 3 67-381.

Kyle, J. G., Kyle, J., & Woll, B. 1988. *Sign language: The study of deaf people and their language*. Cambridge university press.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

Rosen, R. S. 2010. American sign language curricula: A review. *Sign Language Studies*, *10*(3), 348-381.

Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. 2023. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, *15*(7), 5930.

Stokoe Jr, W. C. 2005. Sign language structure: An outline of the visual communication systems of the American deaf. *Journal of deaf studies and deaf education*, *10*(1), 3-37.