Peer Review Report

By - Aditya Singh Thakur

Gesture Recognition Based on Deep Learning

Summary

This project is based on the topic of Gesture Recognition in Computer Science, and is based on computers recognizing human body movements. However, the project focuses only on hand movements. The gestures of numbers have been used for training and testing. The project is based on PyTorch (an open source machine learning library). The Datasets used in the system are Sign Language MNIST and Kinect Leap Dataset. LeNet and ResNet have been used as model network and the performance between the two networks have also been compared. Also, the feasibility of the project has been checked by implementing it on a product using a trained model.

The project clearly states its motivation and various technical advantages such as its application in virtual reality and sign language. The project is of high interest for all audiences as it addresses the issue of gesture recognition that has a high frequency usage in our daily lives and may also play a major role in our future, such as in one of the commonly used electronic devices, iPad, which supports gesture operation, which requires the ability of gesture recognition. Gesture Recognition has a broad range of applications in human-computer interaction. For example, in virtual reality and augmented reality, gesture recognition can be applied widely. Also, in the smart home domain, Gesture Recognition can be utilized to remotely control smart appliances and home robots. It can be conducted with techniques from computer vision and image processing which may open completely new doors for future technologies.

The two datasets used in this system which are Sign Language MNIST and Kinect Leap Dataset have been thoroughly explained. The Models used (LeNet5 and ResNet34) have been depicted accurately and explained in a detailed manner with proper images and data. The Implementation Section has been covered in detail and the setup has been explained clearly in such a manner that the project could be easily reproduced if need. The language used is crisp and easy to understand while following all the depiction of the codes used as well. The results have been shown graphically and precisely depict the variation of accuracy and loss of the system during the Training and Testing Process.

The accuracy of LeNet5 dealing with the MNIST Dataset is 100% in training process and 87% in testing process. But for the Kinect Leap Dataset, even the training accuracy can only reach 55% and the testing accuracy is no more than 40%. Meanwhile for ResNet34, In the Training Process, all accuracy level can reach 100%. And for Test Process ResNet34 can also reach 90% accuracy which is relatively higher than the performance of LeNet5. Thus, showing that **the performance of ResNet34 is much better on both datasets**. LeNet is more suitable for pictures whose resolution are lower than 28 x 28 with faster calculation and less occupation of space. ResNet works better on larger images with deeper model and longer training time.

For the future aspects of the project, the authors can work on various existing methods and techniques to improve ResNet. The project deals with static images, but, however in most cases, it is dynamic recognition that truly takes place. Hence, they could also work on that, furthermore they could also learn how to draw frames from a video and combine several images together to recognize gestures.

Thus, overall, I think the project was executed very well. It successfully finished the training and testing of the two chosen neural networks and compared their performances. It also implemented a final product that can perform as an App to recognize gestures using an algorithm. Also, the fact that the product can automatically chose the best model amongst the two (LeNet and ResNet) just by analyzing the image itself is pretty useful feature and can be extremely helpful with the overall speed and accuracy of execution. Below are some of the nitpicks that I found while examining the paper, they can be addressed and used in making the project even better.

Nitpicks:

In the Abstract and Introduction Section, the R in Gesture Recognition should be capital, also the C and S in Computer Science should also be in capitals. (both nouns)

In the Section 3.1, LeNet5, the word ‘convolution’ has been spelt wrong throughout.

In the Section 4, heading ‘Implementation’ has been spelt incorrectly. In the second line, ‘separate’ has also been written incorrectly. In section 4.1, ResNet34, ‘The first layer **has** 3 residual blocks..’. Also, ‘The fourth layer has 3 residual **blocks..’.**

In Section 4.2, Dataset Adjustments, Sign Language MNIST Dataset should be in capital. (noun) Further, ‘makes us a problem’ could be replaced with ‘creates a problem’. Also, the word ‘layer’ has been spelt incorrectly. The sentence - ‘Meanwhile, LeNet5 can neither learning an image with too large scale..’ is not understandable. Moving on, the word ‘he’ on the last line of page 6 is unnecessary.

In Section 4.3, the heading ‘Training Process’ and the word ‘the’ in the paragraph has been spelt incorrectly.

In Section 5, paragraph ‘section’ is wrongly spelt. The Kinect Leap Dataset should be in capitals. Further in the paragraph, the words imaginably, basically, relatively, maintains, especially, have been spelt incorrectly. Also, grammatical errors, **results are** instead of result are and **low effect** instead of low affect.

In the Application Section, the first paragraph could be changed accordingly for a better understanding. The diagram mentioned below is barely readable and could be enhanced or substituted with a clearer one for better understanding.

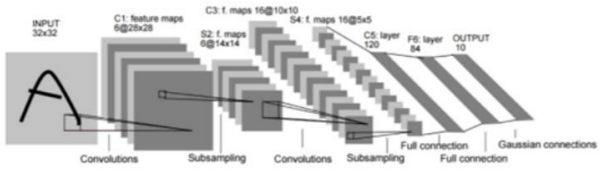

Figure 1: LeNet5