# Peer Review for Project:

# Gesture Recognition Based on Deep Learning

Summary:

Human hand gesture recognition using deep learning methods is a one of the computer vision research topics. In this project, writers implemented two classical neural network, LeNet and ResNet, and trained these networks on two gesture image datasets, Sign Language MNIST and Kinect Leap Dataset. In the result of experiments, ResNet outperforms than LeNet in general. Moreover, writers designed an application product based on their networks, which can automatically choose the suitable model for predication tasks.

High Level Discussion:

Pros:

1. This paper is well-constructed with several academical sections: Introduction, Datasets, Model, Model Implementation, Application and Future Work & Conclusion. It's easy for readers to catch the main ideas and implementation process. The Model and Implementation sections are clear and detailed, which can be reproduced easily.

2. This paper picked up two classical networks LeNet and ResNet, which stand for different network designing philosophies. The tuition is quite meaningful. The comparison of experiment results could be used as reference for those researchers who want to do more exploration on the two kinds of networks. This paper might possibly be a good contrast between relatively shallow network and deep network.

3. This paper compared the results between RGB and gray input images. This result in ResNet-Leap shows that the color has minor impact on gesture recognition, which is reasonable and inspired us when we work on some similar computer vision tasks.

Cons:

1. The novelty of this paper is not that unique. This paper simply compared the training loss and accuracy between two networks. However, more detailed evaluations and explanations would be better if the structures of networks are existing, for example, evaluations on convergence rate or parameters scale, which would also be meaningful for readers. On the other hand, the product this paper implemented should be another novelty, but the description information in section Application is insufficient. It might be the reason that this paper is a draft and would be revised in final version.

2. The experimental setup could be more rigorous. There is more than one factor that cause the difference between ResNet and LeNet: depth of networks, residual blocks and so on, which made it hard to draw the main factors that influence the accuracy of

each network. In addition, this paper chose to resize the images for each network instead of adjusting the input layer for each dataset, which could potentially make it harder to evaluate the functions of components for each network as different networks had different input size. If comparisons could be based on experiments with fewer or one variable, the evaluation would be more reasonable.

3. The design of training process has potential drawbacks. This paper only set one epoch of training for each experiment, while in other deep learning tasks, setting hundreds of epochs are common. In several training experiments, the loss curves didn't converge yet before the end of learning as shown in Figure 4, which would undermine the precision of trained models. It's better to set up larger number of epochs or set up a suitable threshold to stop training.

4. This paper should briefly analyze the characteristics of hand gesture recognition tasks. As this paper focuses on the gesture recognition topic, presenting any reason for why chose LeNet or ResNet for this specific task would be better.

Low Level Discussion Point:

Pros:

1. In Section 4 Implementation, this paper presented core function codes, which clearly helps reader to understanding the details of networks, making this project reproducible.

2. In Section 5 Test Result, this paper explained the experiment result together with a bunch of clear figures. The figures showed the dynamic process of training and testing, providing rich information for readers to learn instead of just providing few numbers as results.

Cons:

1. In Section 2 Dataset, showing some image samples of datasets would help readers to directly distinguish the dataset Sign Language MNIST and original MNIST, as well as the Kinect Leap Dataset.

2. In Section 3 Model, there are too many differences between LeNet5 in figure 1 and description below. For examples, the input of this paper should be 28*28 while in figure 1, it's 32*32. C5 layer in this paper is convolutional layer while in figure 1, it's full connected layer. F6 layer in this paper is cully connected layer with output been 10 while in figure 1 they are Gaussian connected.

3. It's better to add some description for figure 2. For example, what's FLOPs.

4. In 4.2, this paper said 28*28 images are unacceptable for ResNet34 since the images are smaller than a convolutional core. However, as shown in codes above, it seems the kernel size of first convolutional layer is 7. Meanwhile, the statement that LeNet5

can neither learning an image with too large scale need more specific explanation.

5. In section 5 Test Result, some experiment accuracies shown in Figure 4 beyond 1.00, which is impossible in math. That should be typo in y range or bugs in accuracy calculation codes.

6. It should be better to add a section Reference listed in the end of this paper as this paper used several existing figures and some existing network models.

Nitpicks:
For revision convenience, showed in attached PDF files.

Summary:
This project is well-organized from scenarios, models, experiments, analysis, application and conclusion. It implemented two classical networks with rich virtual results but some experiment settings must be revised while more supplement experiments and detailed analysis are encouraged. The part of Application need more content.