ResNet:

The second network we use is ResNet. According to the universal approximation theorem, given enough capacity, we know that a feedforward network with a single layer is sufficient to represent any function. However, the layer might be massive and the network is prone to overfitting the data. Therefore, there is a common trend in the research community that our network architecture needs to go deeper. But when the network goes deeper, the result and may not be so well because a lot of problem raising by the depth of the network. Two kinds of problem, the gradient exploding and vanishing, has been successfully addressed. The ResNet here is to solve the information loss during the process when networks become deeper and deeper. The main idea is to make identity mapping possible. Instead of finding the output, the ResNet focus on the residual part. It uses a shortcut to connect the layers, to feed the original input to the residual output to get the final output answer. Using this method can reduce the information loss during the mapping, because we can get the original input as the output even if the network didn’t work well.

In our project, to make our network become more efficient, we try to apply ResNet. And the result shows that ResNet may work better when the image become larger.

Product:

After training our network, to make the user more convenient to use our network to get the gesture recognition result, we use python to create a product to give the answer.

This product can automatically choose the best model to do the prediction work by analyze the image itself. In the previous part, we find out that different model may have different performance when predict different image. So, after experiment, we set serval thresholds to judge the image and then choose the most suitable model and do prediction.

After choosing the model, we preprocess the image and then feed it the model to get the prediction result, compare with the ground truth to find out whether the output is correct.

For extension, in the future we can apply more powerful architecture to do the gesture recognition. As we all know, the model to get the gesture result from the image may not perform so well to get the hand itself. So, we can use two model to deal with the single image, one is efficient to get the hand result and one is efficient to get the gesture with the previous hands output.