So, I will start introducing the networks we used for our projects. In this project, we use two different neural networks, LeNet and ResNet. I will introduce them one by one.

The first model is LeNet5, which is one of the oldest CNNs. It has been the start point of various networks developed later. There are only 7 layers in the LeNet5, consisting of convolutional layers and fully connected layers. LeNet is able to categorize and recognize patterns using fully connected neural networks without high costs. As mentioned above, one of our datasets is an RGB dataset, but the original LeNet5 is designed for gray scale image. So, we adjust the input channel according to the input dataset and the final output to make the LeNet5 enable to deal with the 3-channel RGB images.

The second one is ResNet, which is a deeper and more complicated model. When the network goes deeper, a lot of problem showed up. Gradient vanishing is one of them. ResNet successfully solve the problem of information loss during the process when networks become deeper and deeper. ResNet adds skip connection in the networks between layers, which makes the network only need to learn the different part between the input and output, simplifying the learning process. Instead of finding the output, the ResNet focus on the residual part. It uses a shortcut to connect the layers, which feeds the original input to the residual output and get the final output. In our design, we first implement a residual block, which adds x to the output H(x) to compute the residual. After that, we can construct the whole network by the superposition of multiple residual blocks.

There are also some aspects we can keep working on for this project. For example, ResNet works better on larger images, but the training speed is relatively lower than the LeNet. In that case, we need to improve the algorithm to work faster. We can try to design and implement different networks that can work faster on large images.

Right now, our project can successfully recognize gestures with static images. However, in daily applications, most cases require the algorithm to work on dynamic recognition. For this part, we still need to learn how to draw frame from a video and combine several images together to recognize gestures.