**Problem Description**

An increasing number of manufacturers rely on robots to stay competitive and solutions where humans and robots work in the same space are entering the market. We wanted to look at how humans could communicate with robots via hand gestures, which eliminates the need for a separate communication interface (most often a control panel). We used two approaches to classify hand-gestures in real-time and were able to get an accuracy of 6.95%

**Data**

- The original EgoGestures dataset has 83-classes of hand gestures, we use 10 classes with about 3,000 samples
- We use 5 frames sampled from videos to evaluate a gesture
- Each frame is resized to 64 x 64, at this size the human eye can still classify the gesture as seen in Figure 1
- 80% train / 10% dev / 10% test split

**Models**

![Model Diagram](image)

**Approach 1: CNN model**

![Graph](image)

**Approach 2: CNN-RNN model**

![Graph](image)

**Remarks**

For the 2D-CNN model we can see from figure 4) that the training error is close to 0% but our evaluated error is over 30% for 11 classes, and for the CNN-RNN model we can see that from figure 5) that the training error is also about 0% whereas the evaluated error is over 30% for only six classes. We thus have a variance problem in both our models, and move so for the CNN-RNN model. This is most likely due to overfitting to our dataset since it is quite small. The error could possibly be reduced by using additional regularization techniques or more data with better resolution but that would require more computational power.