**Title:** Large Scale Activity Recognition in Videos

**Team:** Dewal Gupta

**Abstract:**
Activity recognition in trimmed videos is an established, yet challenging task. The trimmed video classification challenge involves classifying approximately 10 second videos across 400 human action categories. In the Kinetics dataset, the categories include a wide variety of human actions such as applauding, jogging, and kayaking. Each category contains at least 400 video clips for each category, and clips are usually homemade thus including shaky footage, user edited filters on top of the original footage, poor resolution, etc. The top performing models for the 2018 challenge achieved an average error rate of 11%. State of the art networks have moved away from using recurrent networks due to the large scale of data and now use 3D convolutions instead [1]. However, 3D convolutions fall short due to the fact that they don’t force the network to learn temporal relationships in any particular order, and also can be quite computationally expensive. In this research study, the efficiency of LSTMs and other memory networks is evaluated and tested to see if they can achieve state of the art or near state of the art results. One goal of this project is to evaluate whether using multiple LSTM networks operating at different time scales helps the overall network learn better temporal relationships than current state of the art networks. Another goal will be to examine the relationship of 3D convolutions and LSTMs to see if using both can increase classification accuracy. Previous work has included looking at different types of recurrent neural networks such as “ClockNet” [2] and building out infrastructure for loading and preprocessing videos for efficient training.

**References:**