Project Proposal and Feasibility Report

Markus Woodson, Harsha Chivukula, Pratik Prakash, Rishi Khanna

October 9, 2015

1 Project Overview

We plan to create a new engineering solution using a combination of neural network architectures to recognize and classify actions and videos.

2 Project Milestones

- **Milestone #1** - Prior Research - Have a clear understanding of neural networks and foundations of action classification.
- **Milestone #2a** - Create RNN to examine raw video input
- **Milestone #2b** - Create RNN to examine filtered, preprocessed video input
- **Milestone #3** - Combine the different networks
- **Milestone #4** - Collect, analyze, and present results from RNNs

3 Project Schedule

8 Weeks to project demo

Week	Objective
Week 0 (Oct 9 - 23)	Milestone #1
Week 1 (Oct 9-16)	Milestone #1
Week 2 (Oct 16 - 23)	Milestone #2
Week 3 (Oct 23 - 30)	Milestone #2
Week 4 (Oct 30 - Nov 6)	Milestone #2
Week 5 (Nov 6 - 13)	Milestone #2
Week 6 (Nov 13 - 20)	Milestone #3
Week 7 (Nov 20 - 27)	Milestone #3
Week 8 (Nov 27 - Dec 4)	Milestone #4

3.1 Breakdown of Tasks

We will split into two groups of two and split the tasks.

- Milestone #1
  - Learn math required for implementing RNNs
  - Decide on a deep learning framework
  - Research on how to implement an action classifier
  - Research on neural networks and recurrent neural networks
- Milestone #2a
  - Re-read papers and construct RNN according to their architecture

  - Train the network on the type of data specified (video, or custom feature) obtained from the chosen training dataset.
  - Test and determine the optimal hyperparameters

- Milestone #2b
  - Re-read papers and construct RNN according to their architecture
  - Train the network on the type of data specified (video, or custom feature) obtained from the chosen training dataset.
  - Test and determine the optimal hyperparameters

- Milestone #3
  - Create an RNN that takes as inputs the outputs of the two individual RNNs from the previous step.
  - Train the network on the type of data specified (video, or custom feature) obtained from the chosen training dataset.
  - Test and determine the optimal hyperparameters

- Milestone #4
  - Collect and analyze results from all three RNNs
  - Create poster describing the results of the project

4 Research Outline

We know from prior research that RNNs are really good at recognizing sequence data, such as video, speech, time series data[1][3]. In addition, reducing the resolution of the input videos allows us to analyze actions in the videos at a higher level, since the high frequency data is not as relevant to describing action sequences as compared to temporal information[2].

We plan to exploit both of these ideas to build a more sophisticated and robust RNN. We plan to explore if the additional information in the high frequency components will improve accuracy in the classification of actions in videos.

Three of the research papers exploring the ideas of subsampling the videos to reduce resolution and using raw video input are listed in the bibliography.

5 Bibliography

[1] Hierarchical Neural Network Generative Models for Movie Dialogues, Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio

[2] Large-scale Video Classification with Convolutional Neural Networks, Andrej Karpathy, George Toderici, Sanketh Shetty

[3] Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos, Serena Yeung, Olga Russakovsky, Ning Ji