Peter Shagnea and Sarvesh Bhagat

Machine Learning 2 Group Project Proposal

- Problem Statement
    - We have chosen a problem under the domain of action-recognition. We will be implementing a CNN capable of efficiently combining both spatial and temporal information from videos, and performing classification. This has applications in video surveillance, and video storage and retrieval.
- Dataset
    - Selected the '20BN-SOMETHING-SOMETHING-DATASET'. It contains 220,847 videos, divided into a training set of 168,193 videos, a validation set of 24,777 videos, and an unlabeled test set of 27,157 videos. There are a total of 174 class labels, generally in the form of verb - noun - preposition - noun, e.g. "Covering something with something," though they can be more complex. This dataset has adequate samples to train a deep network.
- Deep Network
    - We will implement the top performers from the twentybn leaderboard, and attempt to improve the performance. Negative and positive results will be documented, in order to assess the state-of-the-art action recognition methods.
- Framework
    - We will use Keras or Pytorch. Based on research we will decide which one to use.
        - Keras - We have familiarity with keras from other coursework.
        - As per our research so far, PyTorch has built-in modules for 3D convolution layers and we might use that. It is also the framework used in many of the papers we are basing this project on.
- Reference and Research
    - A variety of resources are openly available in literature. MIT has a recent paper, titled "Temporal Shift Module for Efficient Video Understanding," and a paper from Huaqiao University provides a comprehensive survey of recent methods, "A Comprehensive Survey of Vision-Based Human

  Action Recognition Methods."
- Performance Matrix
    - Top-1 and Top-5 will be used to judge network performance.
- Schedule

Proposal	November 10, 2019
Prepare Environment Prepare Data	November 10, 2019 - Nov 12,2019
Do Research Define/build Convolutional Neural Network	November 13, 2019 - Nov 20,2019
Train, Tune and test Network	November 20, 2019 - Nov 30,2019

Prepare Deliverable and Presentation	November 30, 2019 - December 1, 2019
Final Report	December 2, 2019