Henry Lin, Christopher Vo
henryln1@stanford.edu, cvo9@stanford.edu

CS231N Project Proposal

For our project, we are interested in classifying different dance forms from videos. We plan on breaking these videos into multiple-frame representations that can be processed as 4-D tensor inputs to train deep learning models. Dance is an excellent case study into understanding motion through deep learning due to the unique mix of diversity and similarity between various dances. The broad spectrum of dance allows us to encapsulate the majority of human motion so that understanding and correctly classifying dance forms would be an indicator of the potential for CNN models to understand human motion on a more general level.

There are related works in the field of motion classification, but two of the more relevant ones include a paper on classifying dance motion from Georgia Tech [1] and a paper on motion extraction from videos from CMU [2]. In [2], the general topic of motion extraction from videos using CNNs is introduced while in [1], the comprehensive dance dataset we plan on using is introduced. For the dataset, it is a 1000 video dance dataset with 10 classes for classification. The dataset splits each video into frames, and all together, there are 300,000 frames/images that we will use to train, validate, and test our model. Additionally, we are considering scraping YouTube for more videos if the dataset is too small from a platform by Google Research for extracting YouTube data [3].

There are several deep learning approaches using CNNs previously explored in [1] such as frame-by-frame CNN and skeletal temporal CNN models. In addition to looking at different structures of CNNs, we would like to look at different approaches to representing the data in our model, such as individual frames or stacked images. Finally, we are interested in applying transfer learning from various motion detection models and comparing the performance.

We plan on evaluating our dataset on a test set to determine its accuracy and ability to generalize to unknown dance videos through various metrics such as individual class accuracy, precision, recall, F1 score, etc. Furthermore, we will look at specific cases of misclassification and discuss the possible reasons why.

[1] Castro, Daniel, et al. *Let's Dance: Learning From Online Dance Videos.* Georgia Institute of Technology. 2018. https://arxiv.org/pdf/1801.07388.pdf

[2] Teney, Damien and Martial Hebert. *Learning to Extract Motion from Videos in Convolutional Neural Networks.* The Robotics Institute Carnegie Mellon University. 2016. https://arxiv.org/pdf/1601.07532.pdf

[3] YouTube 8M. *Google Research.* https://research.google.com/youtube8m/explore.htm