

Final Project Proposal

Objective and Dataset

This project will attempt video classification of violence and nonviolence behavior from the RWF2000 Large Scale Video Database for Violence Detection. The objective is to be able to flag violence in video and alert authorities as soon as possible. The database contains raw surveillance videos from YouTube sliced into 5 second intervals at 30 frames per second. This translates to 150 still images per clip. In total this database contains 2000 clips with 300,000 frames (images). The dataset is evenly distributed between violence and nonviolence clips. This dataset contains many different types of clips with a majority having poor image quality due to dark environments, quick movement and low resolutions.

Network

The frameworks may diverse based on different models from group members, but they are mainly focused on Pytorch and Keras. For deep networks, there are many methods that have been used for video classification and thankfully most are well documented. To start we will divide our network into optical and flows and RGB for inputs into the model. The first attempt will be to implement a 3D Convolution Neural Network as it has been proven to perform well when classifying human action from video. Unlike Conv 2D a Conv 3D when applied to video analysis can capture the motion information between multiple continuous frames. After establishing a basic C3D we can explore adding iDT and linear classifier sigmoid function to improve accuracy. Once our model is functioning properly, we will move towards more complicated networks like Long-term Recurrent Convolutional Networks (LCRN) which builds upon RNNs and LSTMs. Two Stream Fusion and Temporal Segment Networks can also be explored if the networks already mentioned are functional. Network performance will be judged by accuracy of the binary classification.

Reference Materials

The main reference material that will be used is Ming Cheng's GitHub account found at: <https://github.com/mchengny/RWF2000-Video-Database-for-Violence-Detection>. This is a well-documented process from downloading the dataset, preprocessing steps and a CONV3D network using both optical flow and RGB. We will be fully utilizing the documented preprocessing script which converts the clips into .npy files (which is a tensor) and drawing from the built network as we create our own custom network.

Schedule

November 15, 2020 by end of Day:

Must Do: Download the Dataset and perform basic EDA

Goal: Implement a simple CONV3D Model (or implement model provided on github)

November 22, 2020 (Exam Week) by end of Day:

Goal: Implement a simple CONV3D Model (or implement model provided on github)

November 29, 2020 by end of Day:

Must Do: Have a full CONV3D Model implemented and producing results. Begin writing and obtaining graphics for the final report

Goal: Hyperparameter tune CONV3D Model. Begin Exploring other frameworks and networks (LCRN, TSN)

December 6, 2020 by end of Day:

Must Do: Finish the Final Report