CS 231N Project Proposal
Complex Video Querying with Weak Supervision

Swetha Revanur
Stanford University
srevanur@cs.stanford.edu

Vishnu Sarukkai
Stanford University
sarukkai@stanford.edu

As many fields have become more data-rich, acquiring labeled training data has become the primary bottleneck in supervised machine learning. This motivates the need for an alternative, more flexible supervision strategy (i.e. weak supervision) that employs lower-quality or noisier labels. We seek to build a model to automatically detect complex actions in videos, such as "car turning left" using weak supervision. This will allow us to create general models for various complex video queries without the need for a benchmark dataset. To provide context and background, we plan to refer to major papers by Ye [3], Wang et al. [2], and Ahsan et al. [1]. Specifically, Ye discusses automatic event discovery in videos, while Wang et al. extends this to untrimmed videos. Ahsan et al. applies deep convolutional GANs with a weak supervision approach to actually label videos. Using publicly available video datasets like THUMOS14, ActivityNet, and UCF101 (all used in the papers mentioned above), we aim to combine the methodologies presented by Wang et al. (using CNNs for video querying) and Ahsan et al. (using GANs to enable weak supervision) into an end-to-end model. We also want to extend their work to handle more complex actions, while achieving comparable accuracy. We will present these results in a paper including a variety of plots and figures to illustrate our work.

We will be working under the guidance of Paroma Varma from the Stanford InfoLab.

References

[1] U. Ahsan. Discrimnet: Semi-supervised action recognition from videos using generative adversarial networks, 2018.

[2] L. Wang. Untrimmednets for weakly supervised action recognition and detection, 2017.

[3] G. Ye. Large-scale video event detection, 2015.