Action Recognition

Truman Thames – tthames@broncos.uncfsu.edu
Mentors : Dr. Boqing Gong bgong@crcv.ucf.edu
Lijun Lilil ijun1990@buaa.edu.cn

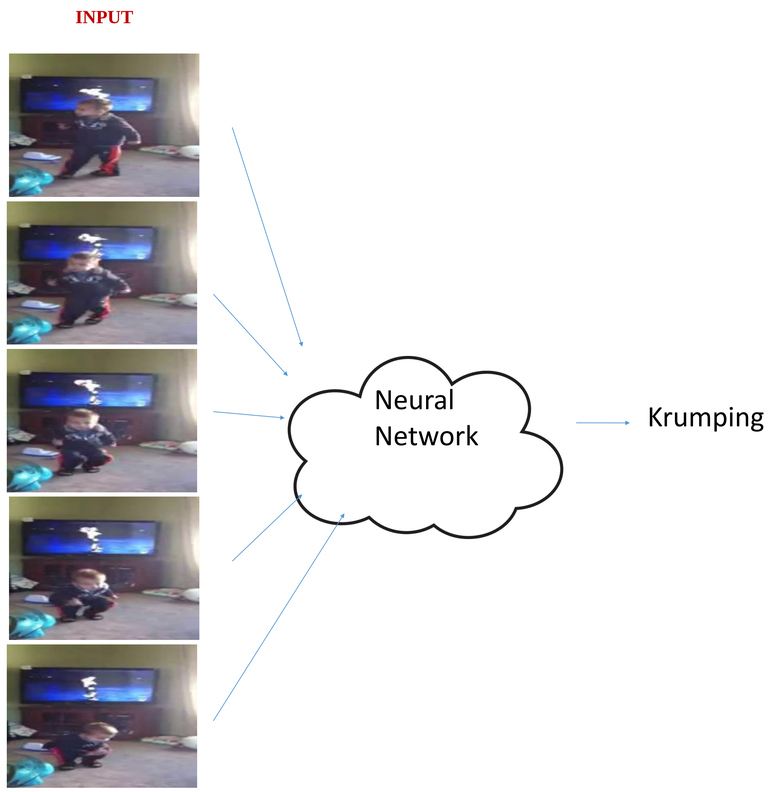
UCF | Center for Research in Computer Vision
UNIVERSITY OF CENTRAL FLORIDA

INTRODUCTION and OVERVIEW

Action recognition is the field of computer vision in which computers are tasked with classifying the primary action in a scene. In general the scene can be a video or a picture. When the scene is a picture the problem can be called a subset of image classification.

OBJECTIVES

The objectives of this project are to improve upon the state of the art in video action recognition using a neural network trained from the ground up. We also aimed to learn more about the effects of vastly different data on similar networks.



KEY PROBLEMS

One of the primary problems in working with videos for action recognition is the amount of data that must be pre-annotated and trained on to have a successful algorithm. Recently, a new dataset, “kinetics”, has appeared on the scene. It has more than ten times as many videos as the most common datasets UCF101 and HMDB51. This dataset should help solve the main problem of lack of data in video action recognition.

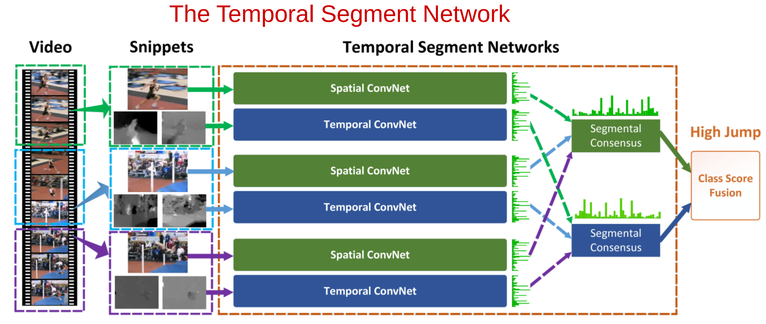
DATASET and METHODS

The dataset that we are most interested in using for this project is the kinetics dataset. This dataset contains more than 260,000 short links to youtube videos that are each classified into one of 400 categories.
The model I use as a base is the temporal segment network created by Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, Luc Van Gool.

Choose Dataset	Process Dataset	Train Model
• Robust (400 classes, 40 videos per class) • New (released Spring 2017)	• Extract RGB frames • Extract Optical Flow • Create index file for frames	• Modify model for use with kinetics dataset

CONCLUSIONS

My conclusions thus far have been inconclusive. I need to further optimize the model and the training.



ACKNOWLEDGEMENTS

I would like to acknowledge Dr Gong, Lijun Li, Dr. Shaw, Dr. Lobo and the CRCV REU program grant award no. 1461121 for the opportunity to research in the area of computer vision.

REFERENCES

http://yjxiong.me/papers/eccv16tsn.pdf
https://github.com/yjxiong/temporal-segment-networks
https://arxiv.org/pdf/1705.06950.pdf

FUTURE WORK

In the future I plan to expand preprocessing done on the dataset and include the optical flow for the rgb frames. I was not able to do so in this project as I would have not had enough time to complete the optical flow extraction.
I also plan to