Libby Zorn
Brown University Computer Science Sc.B. 2016 Capstone Abstract

Title: **Video Classification with Neural Networks**
Faculty Sponsor: Erik Sudderth

Abstract:

My capstone project was done in the class CLPS1520: Computational Vision, and was done in a group with Eli Rosenthal. As a background, the project included literature review of current research in the use of neural networks to do video classification, and we read three papers:

- Learning Spatiotemporal Features with 3D Convolutional Networks (Tran et al. 2015)
- Efficient feature extraction, encoding, and classification for action recognition (Kantorov and Laptev 2014)
- Dense, Accurate Optical Flow Estimation with Piecewise Parametric Model (Yang and Li 2015)

Our background research led us to two questions to answer in the project: How effective is the classification of videos using neural networks, and what improvements can we make to a baseline classification of just the individual frames of the videos? Our dataset consisted of 14 1-hour-long labeled videos at 30fps of mice from the Serre Lab, with 13 different classes of mouse movement. We first looked at a biological concept called optical flow, which is the change in light perceived by the eye caused by movement in the environment, and found a computational equivalent which we used to generate motion data from the videos. This computational concept, called mpegflow, generates a series of motion vectors used in video compression. We trained and tested our neural net using the motion vectors from mpegflow in hopes of increasing classification accuracy.

My focus on the project was to configure and run Alexnet (the model of neural net we used) on both still images for our baseline training and the motion vectors. This included getting the labeled videos from the lab and taking the motion vectors Eli created by running mpegflow and converting them into lmbd format to run on Alexnet. Then, I read the documentation on Alexnet in order to configure that data we had on the machines we were using and helped Eli set it up. Finally, I trained and tested both datasets on the Alexnet.

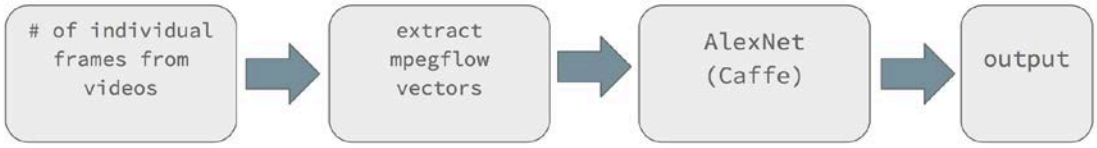
We found that the accuracy decreased significantly when classifying based on the the motion vectors, compared with our baseline of still images. These were not the results we expected, but it turns out that the motion vectors contain very little information when the mice are not moving, and so for example if a mouse is performing the action “hang” or “rest” and is not moving very much during the action, the motion vectors will be almost identical. Also, the still images were 256 x 256 while the motion vectors were only 30 x 40, we were classifying with much less information when using the motion vectors. Future explorations of this project would be to use spatiotemporal filters, which are a more robust model of optical flow, or to use recurrent neural nets to learn long-term temporal features.

System Diagrams:

Baseline training / testing

- each image is 256x256
- train: ~115,000 images
- test: ~28,000 images
- randomized 80/20 split

mpegflow training / testing



- motion vectors are 30x40
- train: ~115,000 images
- test: ~28,000 images
- same 80/20 split as before
- Training is roughly 20 times faster than video frames.