Journal Report 8
11/11/19-11/18/19
Avyuk Dixit
Computer Systems Research Lab
Period 4, White

Daily Log

Monday October 28

On Monday, Saketh and I spent time discussing how our model architectures interact and could be used together to identify instances of violence from surveillance feed. Saketh's model identifies emotions from static frames whereas my new architecture marks violent behavior using a sequence of frames. The easy solution to this is to just have a set number of frames over which the pose is determined to be violent of non-violent and the majority emotion (binary - violent or non-violent) in those frames dictates whether or not there is violence. Another question we thought about was how exactly the model would make a final decision. Though it could give a probability of whether or not violent behavior exists, neither of our architectures are fit to do that. We will probably continue thinking about how to resolve this issue next week.

Tuesday October 29

I was not at school on Tuesday.

Thursday October 31

I cleaned up my code so that only the parts that worked were left. Joined GitHub classroom, found my repository, and pushed all my code there. Helped Saketh do the same. Looked a little bit more into Convolutional LSTMs and tried to best understand the math behind them. It was a little bit hard to do since there's not much work on it right now and the original paper isn't especially great. I had to leave around 45 minutes early on Thursday.

Timeline

Date	Goal	Met
Today minus 2 weeks	Review OpenPose code and determine best statistical representation for poses.	Completed
Today minus 1 week	Decide on architecture to use	Finished
Today	Run OpenPose code through dataset of frames marked as violent or non-violent.	
Today plus 1 week	Create model, get it ready for use	
Today plus 2 weeks	Run model, refine it.	
WINTER GOAL	Achieve 70% accuracy	

Reflection

I think I'm at a good point now to start writing code for the model and developing it. The only things I'm still unsure about are 1] how to account for bystanders and 2] how to determine changes in a person's pose over time. For example, if a person wasn't in frame 1, part of their body appears in frame 2, and all of their body appears in frame 3, will that series of frames by an accurate representation of violent behavior. What would stop the network from learning the wrong relationships in cases like that. Maybe only frames that are marked violent can be used as sequences and transition frames from non-violent to violent can't be. If that were the case though, it avoids an important real-world scenario.