Project Evaluation Report
Review Id: R16

Description of the Problem:

This project tried to predict the controller input for Super Smash Brothers game using the video frames that was collected while playing the game. Now as there was no data available the major part of this project was collecting data. The architectures used were modified CNN (as CNN has proved to be quite successful in image processing tasks) like AlexNet, SlimVGG and LeNet and it was found that AlexNet gave the best result, also as data was self-created so it was not a lot, so it was important to avoid overfitting of the data and so dropout techniques were used for regularization and it achieved better results than without dropout as was expected.

Positive Aspects of the Analysis:

The positive aspect of the project was the way the data was prepared by recording gameplay frames using an off-the-shelf screen capture tool, they used k-nearest neighbors to find the largest cluster of pixels close to the character's main color and mark the center of that location as the location of the character, also they used data augmentation to artificially inflate the dataset by augmenting the next batch of observation on the CPU while training with the current batch on the GPU, this helped them to generate large amount of augmented data without storing them on the disk and without slowing down the training which was very innovative.

Negative Aspect of the Analysis:

The data collection techniques used were very innovative but could have been improved by using Neural Networks to learn the features from the frames rather than just using k-nearest logic as it might miss out on some important features.

Possible Extensions:

As pointed out in the negative aspect of the analysis what I can suggest for future work on this project is to try detecting the features from the frames using neural network(Representational Learning).Also we can use LSTM in two ways to improve the model , the first is to use LSTM for data augmentation as the collected frames must have similarity between them and so LSTM can be used to generate more frames given some already created frames , and so we will have more data at our disposal to train our model and which will lead to better predictive powers for the model. Also as LSTM helps in finding relations between nearby frames and also frames that are far apart, this can potentially also help in improving the predication of the controller input which will improve the predictive power of the model, so a modified architecture using CNN stacked on top of LSTM might improve the model performance further.