A CONVOLUTIONAL APPROACH TO ACTION ANTICIPATION

SHELLY SRIVASTAVA

190385633

MSC IN ARTIFICIAL INTELLIGENCE WITH INDUSTRIAL EXPERIENCE

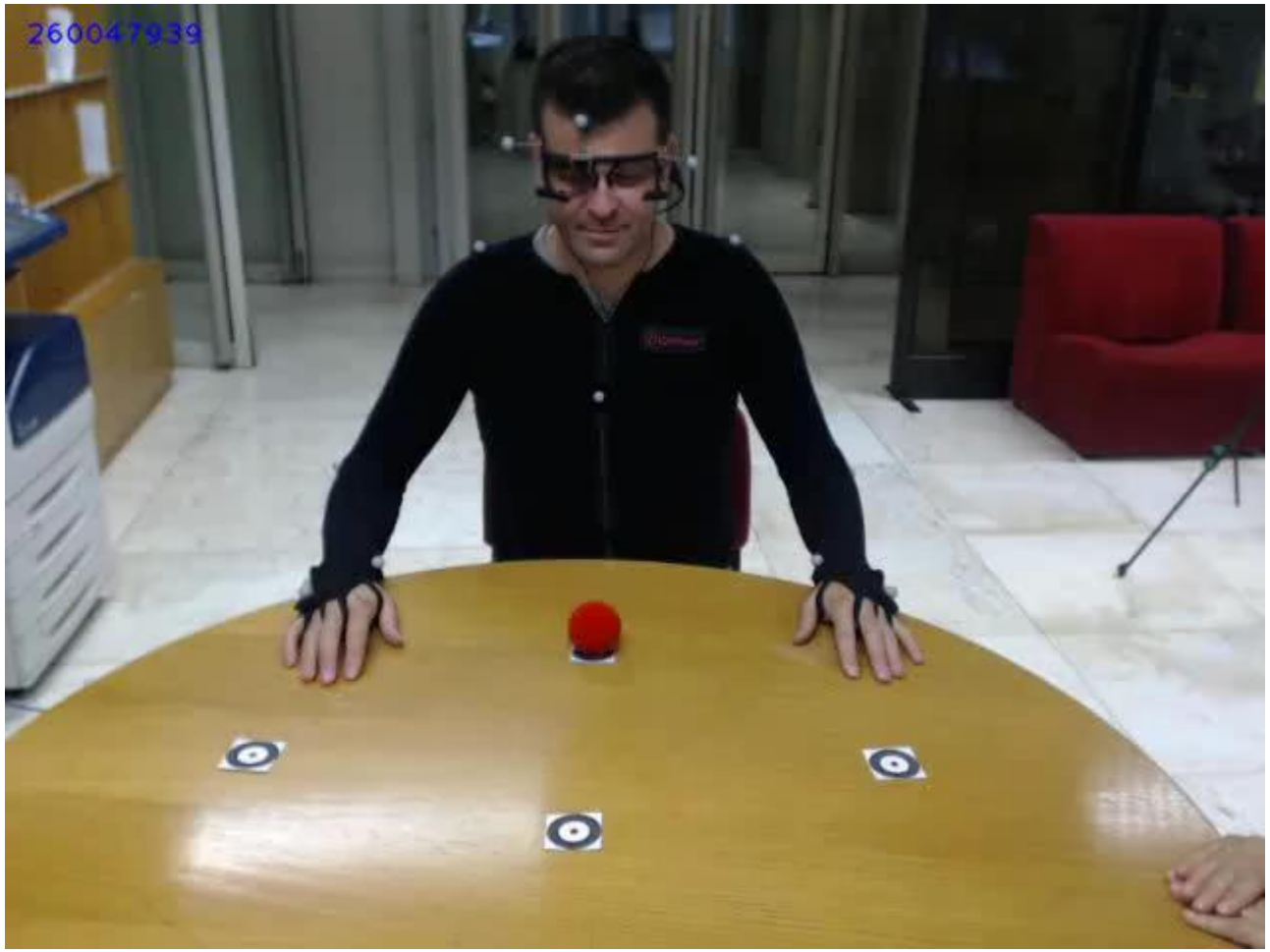
Code available at https://github.com/ShellySrivastava/Action_Anticipation

AGENDA

- What is action anticipation?
- Problem Statement and Dataset
- Video Presentation
- Previous Work
- Implementation
- Performance Comparison
- Future Work

WHAT IS ACTION ANTICIPATION?

Video Source: https://www.youtube.com/watch?v=SdI6lrQUa8s&ab_channel=KUKA-Robots%26Automation



PROBLEM STATEMENT AND DATASET

- Anticipation model predicts the action class based on partial/incomplete action sequences
- Acticipate Dataset is used for the experiments
- It has 12 action classes

WORKING

PREVIOUS WORK: SANTOS ET AL. (2019)

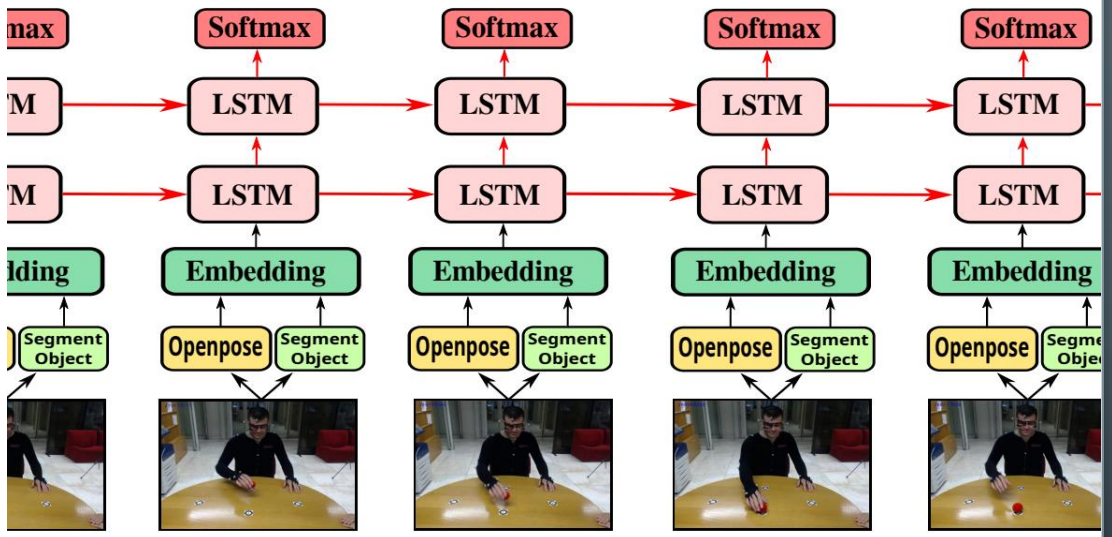
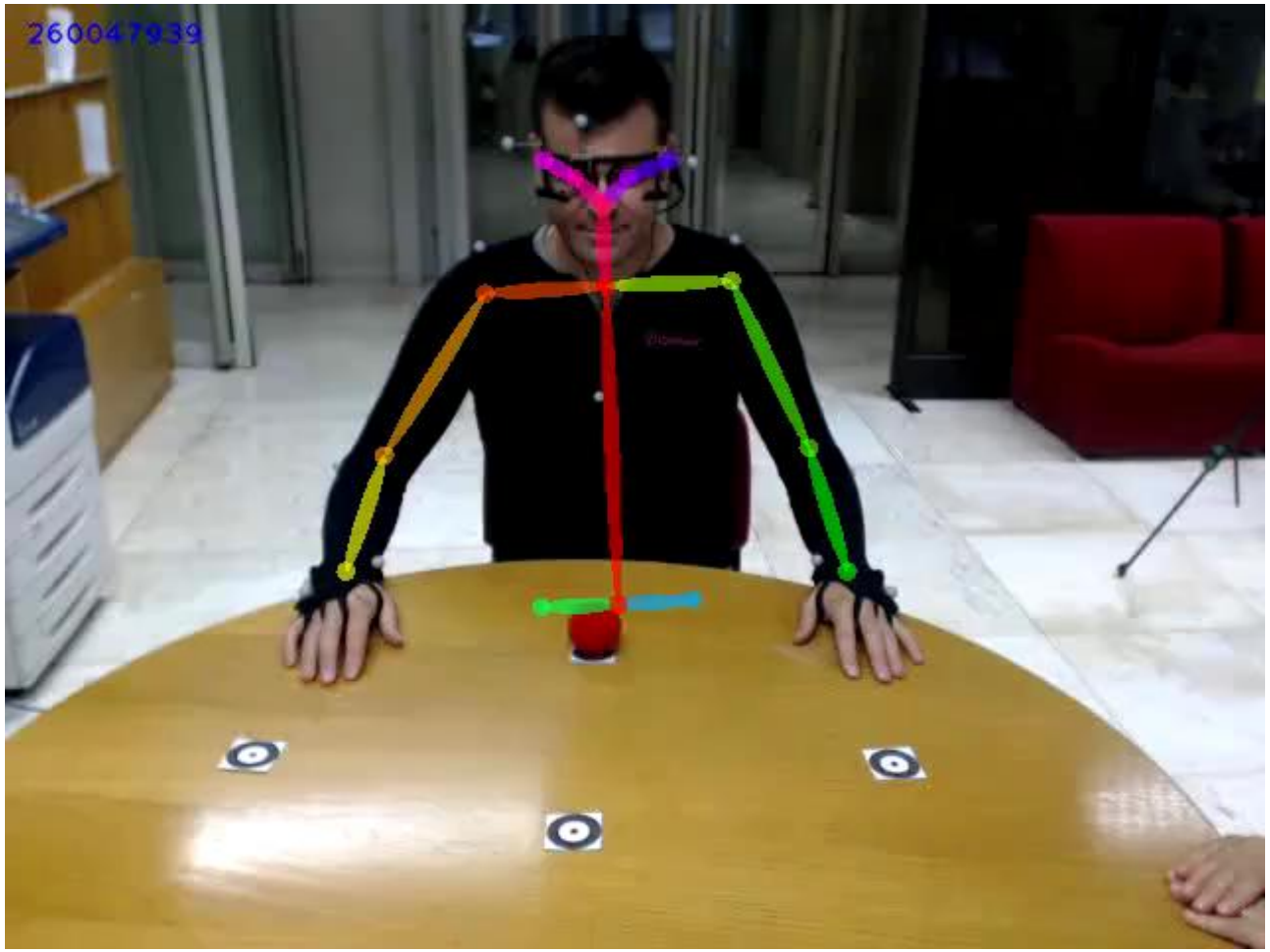


Figure 7: Proposed model architecture

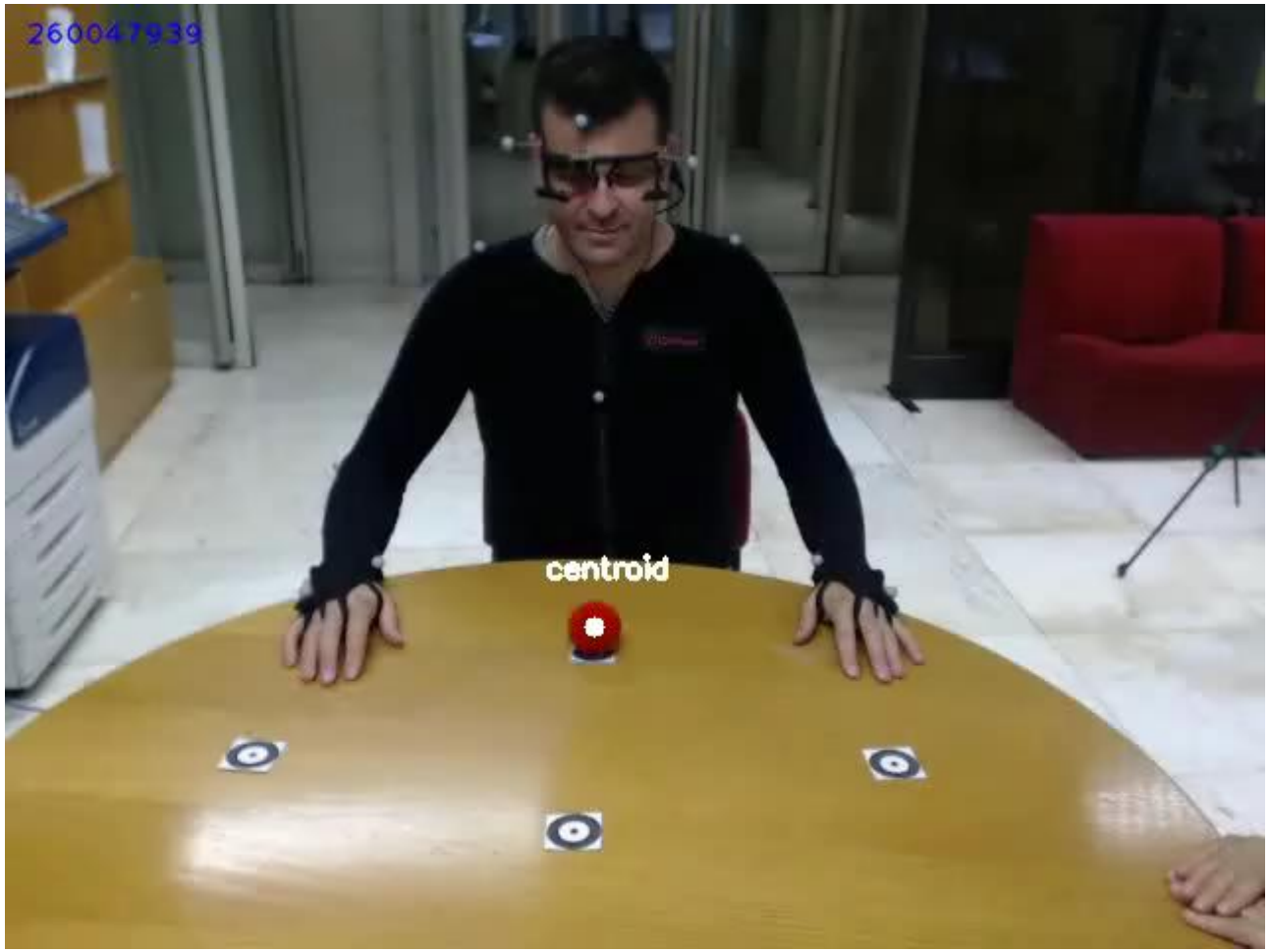
- LSTM based architecture
- Used limb movements with contextual information such as head positions and ball positions
- Their pre-processing included OpenPose and OpenCV

MY IMPLEMENTATION



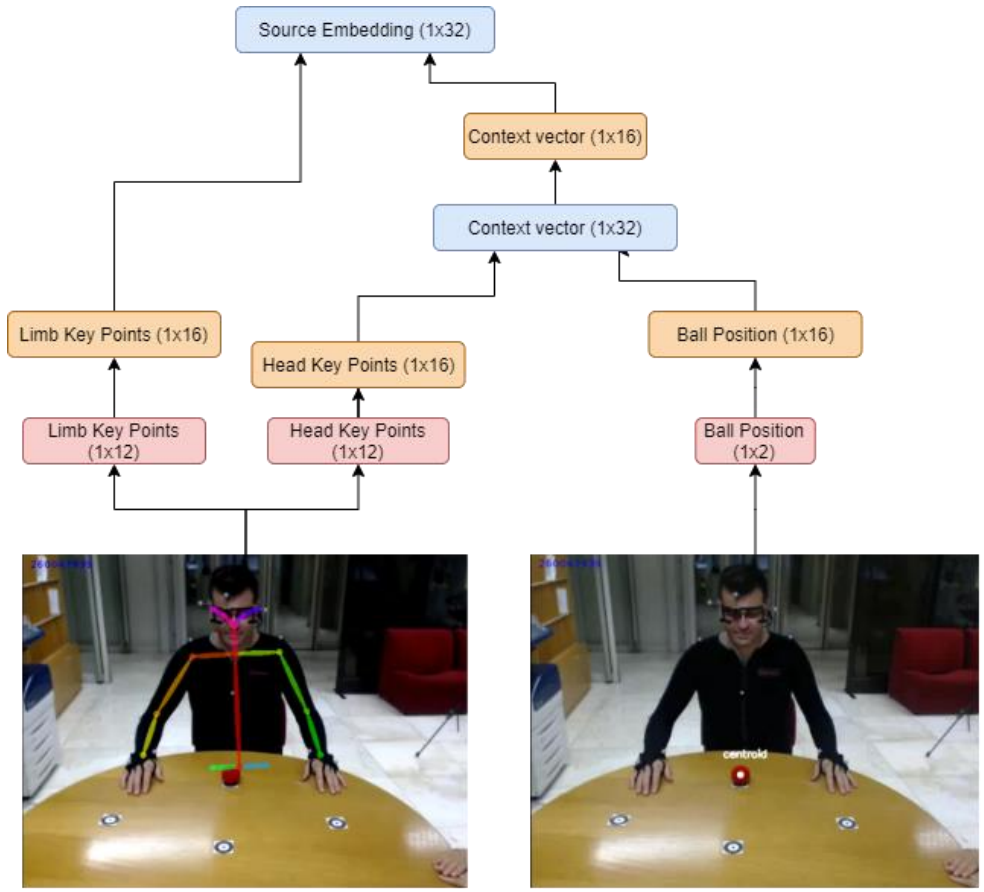
PRE-PROCESSING

- OpenPose
- Get skeletal joint positions of the actor



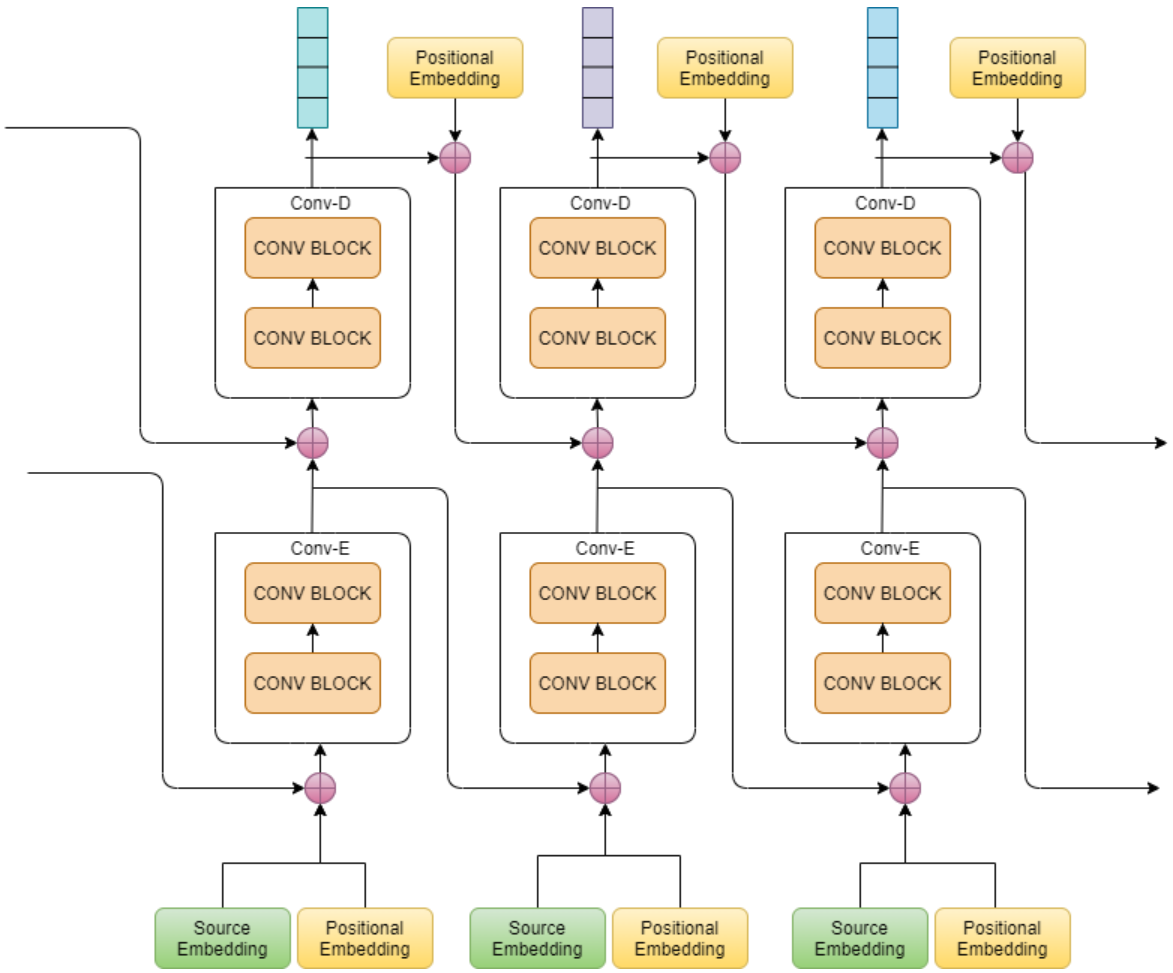
PRE-PROCESSING

- OpenCV
- Get position of ball in the frame using segmentation



PRE-PROCESSING

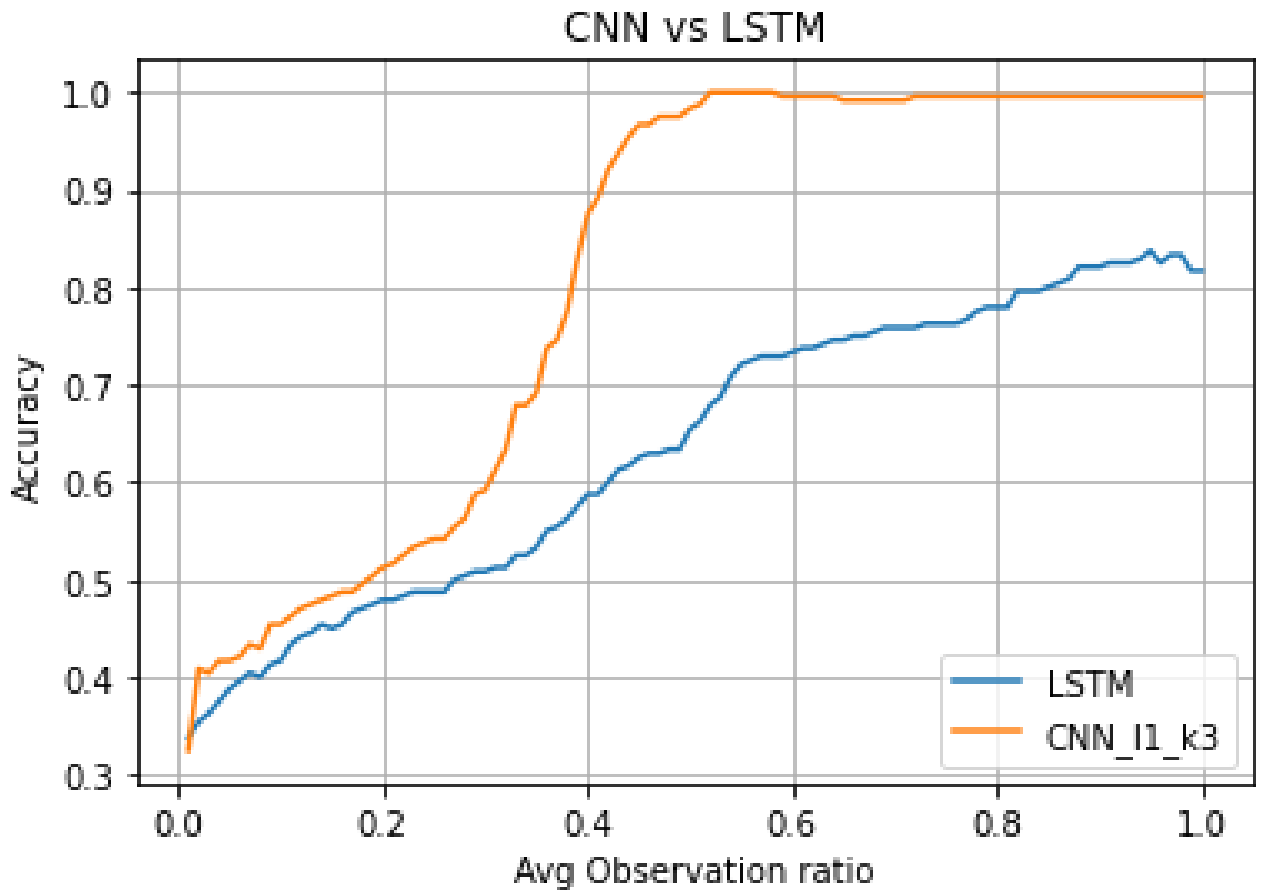
- Pre-processing steps are aligned with Santos et al. (2019)
- Limb movements + context (50-50)



PROPOSED CNN MODEL

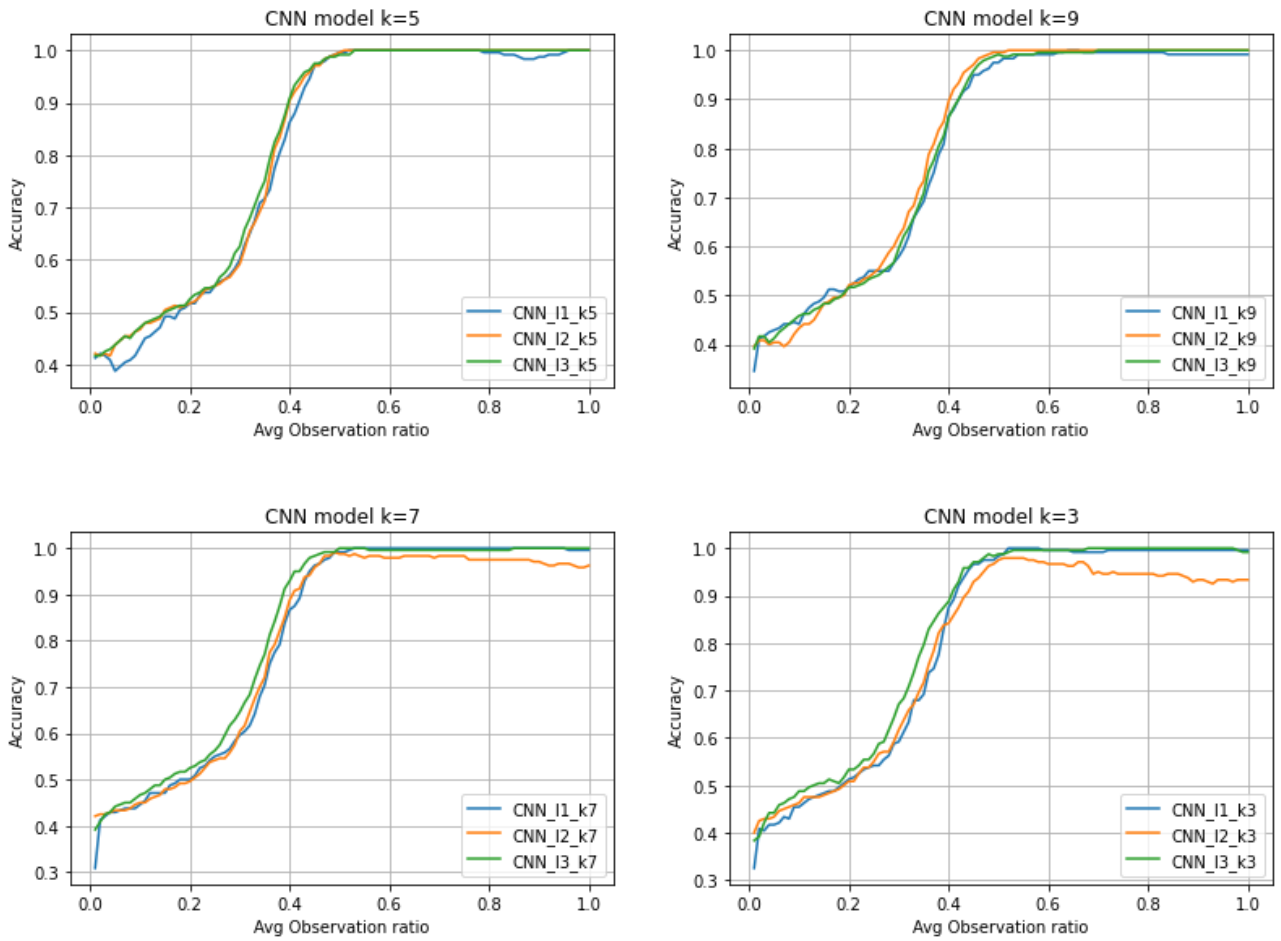
- Uses CNN instead of LSTM
- Convolutional blocks are inspired by Gehring et al. (2017)
- Convolution blocks uses GLU activation
- Uses positional embedding
- Uses feedback loop

PERFORMANCE COMPARISON



PERFORMANCE

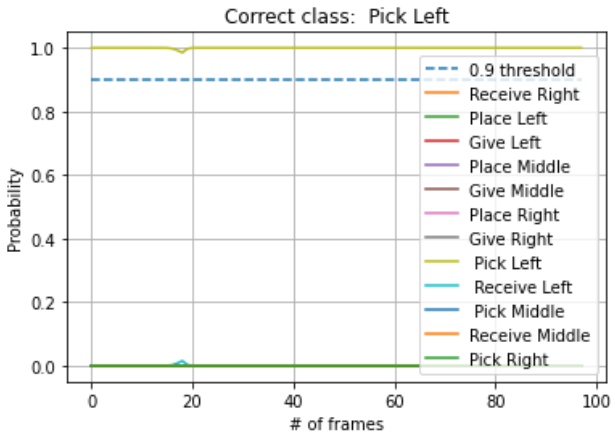
- Base LSTM vs CNN
- CNN model is better at action recognition
- CNN model has better anticipation accuracy wrt observation ratio

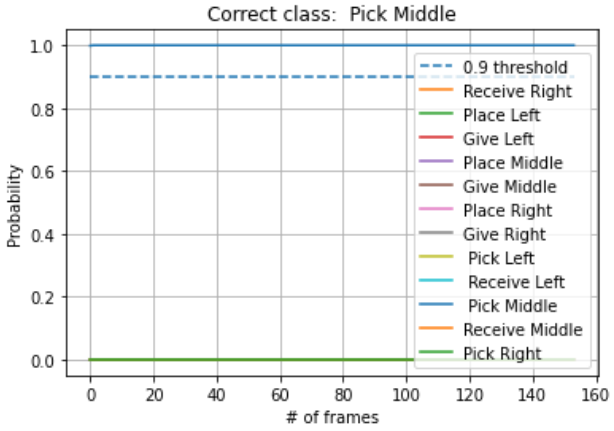


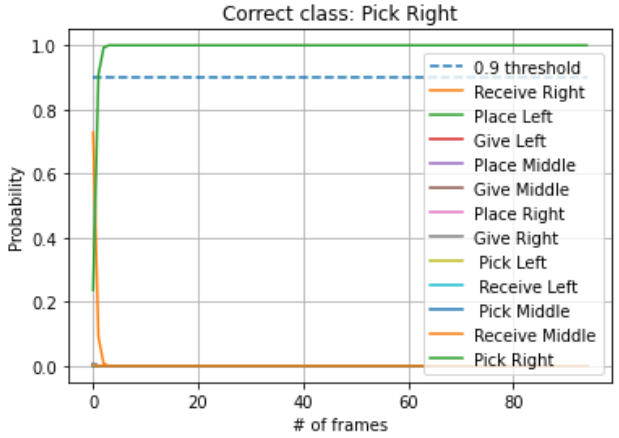
EFFECT OF CONV LAYERS

- Increasing the layers of convolution increases the anticipation accuracy at lower observation ratio
- An anticipation model should correctly predict the action class at lower observation ratio

ANTICIPATION FOR DIFFERENT ACTIONS

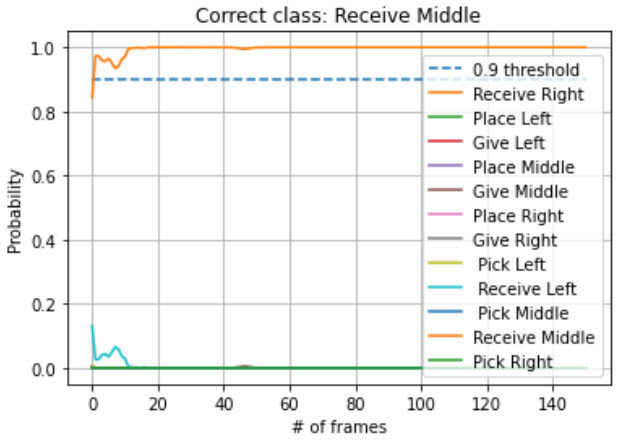


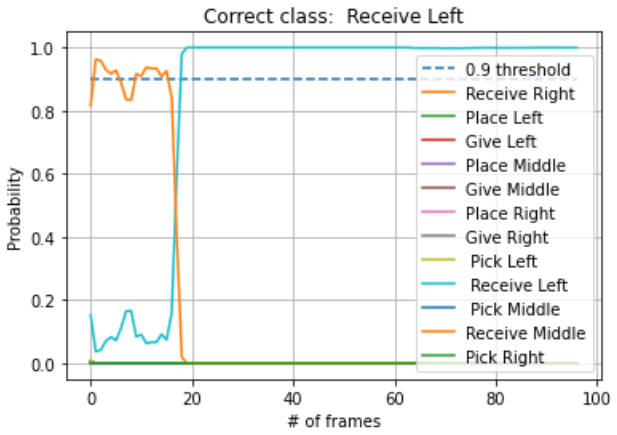


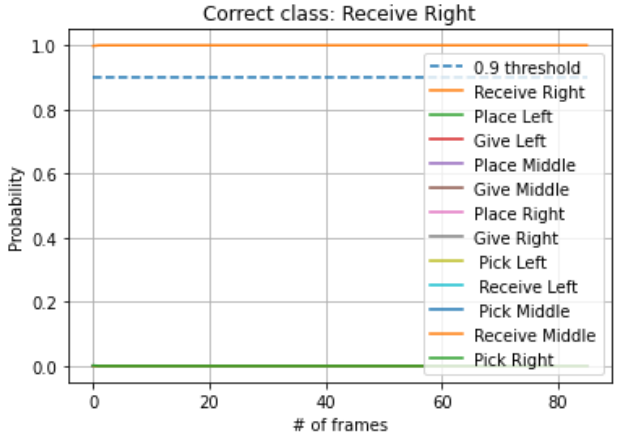


PICK ACTIONS

- Can be anticipated with high probability at the starting of the action sequence
- The model can learn contextual information which is provided

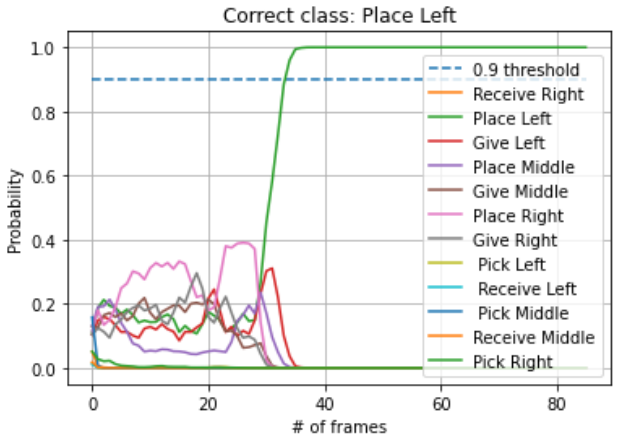


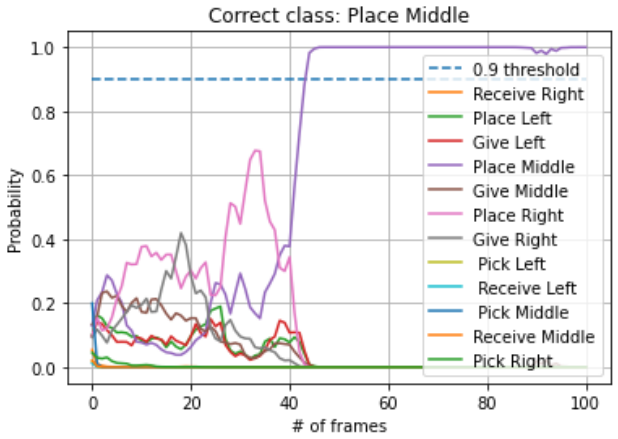


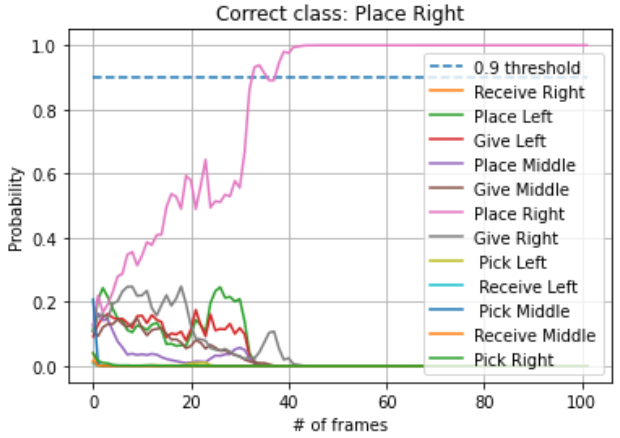


RECEIVE ACTIONS

- Learns contextual information
- Also makes error due to overconfidence: Receive Left

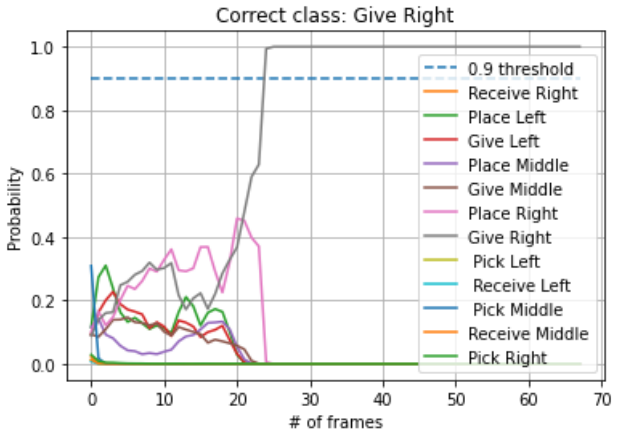


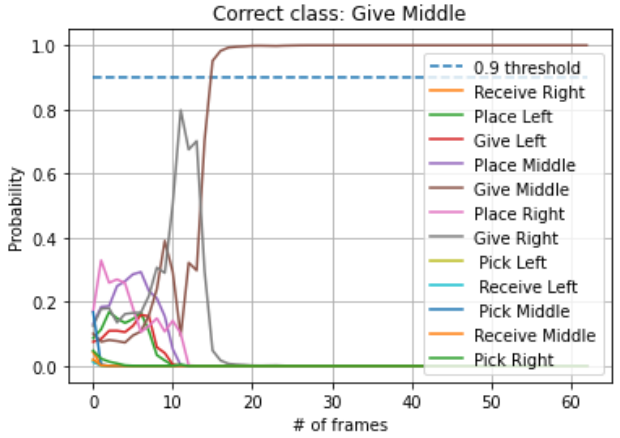




PLACE ACTIONS

- The confidence of the action class increases with passing frames





GIVE ACTIONS

- The confidence of the action class increases with passing frames

FUTURE WORK

- Stochastic CNN model for Action Anticipation – Bayesian Neural Network

QUESTIONS ?

THANK YOU