

LSTM: Simpler architecture for Activity Recognition in Videos

Introduction:

The machine learning models at inception have performed image recognition on a series of images to track the motion using CNN and predict what the activity is. The performance is poor compared to the modern models. Currently there are models using CNN, R-CNN, RNN, CNN-LSTM networks to perform activity recognition however, they are complex and requires a huge load of computation. The CNN is generally used to extract features on video stream using OpenCV library. This project is not a replication but based on a research paper by Anish Lakapragada (Research student at Stanford University).

Idea Proposal:

The basic idea of this project is to develop a simplified activity recognition model compared to existing models. The existing models take the video in contrast to the mobile-optimized MobileNetV2 convolutional neural network architecture with over 3.5 million parameters and need umpteen amount of data to train and perform the prediction. This project focuses on usage of LSTM and Mediapipe to enhance the accuracy and efficiency of activity prediction on live stream video. This is an attempt to reduce the amount of data to train the NN model by maintaining the complexity of the Neural network comparatively simple than existing models by using just half a million parameters.

Description:

To extract the coordinates, we would use MediaPipe, a framework by Google that detects the landmarks on a person's face, hands, and body. We take the first couple of frames of a live video and for each frame, we concatenate the feature vectors and input them into each timestep of an LSTM model. We experiment with subsets of landmarks provided by MediaPipe. The LSTM Neural network is trained using this data, hence, increasing the efficiency of computation and prediction.

Bibliography

R. Girdhar, G. Gkioxari, L. Torresani, M. Paluri, and D. Tran, "Detect-and-Track: Efficient Pose Estimation in Videos," in CVPR, 2018.

Z. Zhao, Z. Zhu, X. Zhang, H. Tang, J. Xing, X. Hu, J. Lu, and X. Qu, "Identifying autism with head movement features by implementing machine learning algorithms," Journal of Autism and Developmental Disorders, pp. 1–12, 2021.

Team: Ishani Ghose, Harish Udhaya Kumar

Course Instructor: Prof. Ahmed Elgammal

E. Chong, E. Clark-Whitney, A. Southerland, E. Stubbs, C. Miller, E. L. Ajodan, M. R. Silverman, C. Lord, A. Rozga, R. M. Jones, et al., "Detection of eye contact with deep neural networks is as accurate as human experts," *Nature communications*, vol. 11, no. 1, pp. 1–10, 2020.

H. Kalantarian, P. Washington, J. Schwartz, J. Daniels, N. Haber, and D. P. Wall, "Guess what?," *Journal of healthcare informatics research*, vol. 3, no. 1, pp. 43–66, 2019.

A. Estes, J. Munson, S. J. Rogers, J. Greenson, J. Winter, and G. Dawson, "Longterm outcomes of early intervention in 6-year-old children with autism spectrum disorder," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 54, no. 7, pp. 580–587, 2015.