Human Activity Recognition Using Machine Learning Techniques

Mr. Lakshmisha S Krishna [1], Ankita Yeleswarapu [2], Jeevan R [3], Mohammed Zaid Khan [4],

[1]Computer Science and Technology (AI and ML), Presidency University, Bangalore
[2] Computer Science and Technology (AI and ML), Presidency University, Bangalore
[3] Computer Science and Technology (AI and ML), Presidency University, Bangalore
[4] Computer Science and Technology (AI and ML), Presidency University, Bangalore

Abstract - Human activity recognition (HAR) has a broad spectrum and plays an essential role in applications in a wide range of disciplines. External and wearable sensors are used to recognize human activities. This paper examines the effectiveness of Convolutional Neural Networks (CNNs) in HAR, using the UCF101 dataset—a vast collection spanning different human activities in the form of video clips. The study encompasses the design, training, and testing of CNN architectures optimized for video-based activity recognition. Utilizing transfer learning and stochastic modeling, the CNNs perform well in capturing complex trends and temporal dependencies in the UCF101 dataset. The framework's adaptability to various settings, incorporation of interpretability characteristics, and ongoing improvement via community contributions demonstrate its efficacy for real-world applications. The findings contribute to the progress of CNN-based HAR techniques, stressing their importance in understanding and interpreting human behaviors in convoluted video sequences.

1. INTRODUCTION

Human Activity Recognition (HAR) has emerged as a pivotal research area in various fields ranging from technical to medical. The ability to automatically identify and interpret human activities from sensor data holds tremendous potential across a multitude of applications, ranging from healthcare and assistive technologies to security, smart environments, and beyond. As our world becomes increasingly connected and data-rich, the demand for systems capable of understanding and responding to human behavior has grown exponentially. Convolutional neural networks bring a fresh perspective to human activity recognition by reinventing how systems extract meaningful signals and understand patterns within motion data. Rather than relying on manually defined features, these networks independently discover the informative components hidden inside raw sensor readings. In doing so, they develop a nuanced awareness of the subtle characteristics that distinguish one activity from another, much like how people intuitively parse human movement. The convolutional filters specifically allow the networks to pinpoint localized motifs within the data, making them attuned to fine variations in movement that may escape traditional analyses. Additionally, the spatial positions of the informative features matter less to these networks - they care more about the sequential flow. This means small changes in where someone places a sensor on their body doesn't obstruct the network's capacity to make sense of the data. Their flexibility and scalability primes them to greatly advance activity recognition. In this paper, we’ve used UC101 dataset for training and testing the data. UCF101's diverse and realistic dataset from YouTube videos is a valuable asset for Human Activity Recognition (HAR) research. Considering its large-scale collection, emphasis on temporal dynamics, and inclusion numerous activities, it provides a good a foundation for training and testing robust HAR models. The dataset's open-source nature and accessibility stimulate further interdependence, accelerating enhancements in HAR approaches with practical applications.

2. Literature Review

Sl No.	Author Name and Year	Title	Findings
1	H. Yang, X. Wen, Y. Geng, Y. Wang, X. Wang and C. Lu,(2022)	A Multi-Position Joint Angle Dataset for Human Activity Recognition Using Wearable Sensors	Methods that reliably recognize a variety of human actions, underscoring the potential of wearable sensor technologies for reliable activity.
2	F. Zhou, R. Wang, H. Su and S. Xu(2022)	A Human Activity Recognition Model Based on Wearable Sensor	The combination of CNN and RNN architectures enhances the model's ability to recognize and classify specific actions in the dataset.
3	I. Stolovas, S. Suárez, D. Pereyra, F. De Izaguirre and V. Cabrera(2021)	Human activity recognition using machine learning techniques in a low-resource embedded system,	To successfully forecast activities from accelerometer data, the system makes use of statistical properties, dimensionality reduction (Linear Discriminant Analysis), and SVM classification.
4	Q. Jian, S. Guo, P. Chen, P. Wu and G. Cui(2021)	A Robust Real-time Human Activity Recognition method Based on Attention-Augmented GRU	AASC surpasses traditional GRU models in terms of accuracy and robustness by leveraging attention processes to facilitate efficient temporal connection learning .
5	M. Atikuzzaman, T. R. Rahman, E. Wazed, M. P. Hossain and M. Z. Islam(2020)	Human Activity Recognition System from Different Poses with CNN,	Obtained excellent scores for pose and activity detection accuracy. The stated processing speed is significant since it suggests real-time application potential.
6	N. Amin Choudhury, S. Moulik and S. Choudhury(2020)	Cloud-based Real-time and Remote Human Activity Recognition System using Wearable Sensors	Could correctly classify different human activities at a rate of nine, which is especially useful when taking wearables and cloud computing .
7	R. Saini and V. Maan(2020)	Human Activity and Gesture Recognition: A Review,	CNNs have demonstrated success in accurately recognizing a variety of human actions.
8	M. S. H. Bhuiyan, N. S. Patwary, P. K. Saha and M. T. Hossain(2020)	Sensor-Based Human Activity Recognition: A Comparative Study of Machine Learning Techniques,	Combining (PCA) with (RF) on phone accelerometer data produced the best results in terms of HAR accuracy.
9	L. Xie, J. Tian, G. Ding and Q. Zhao(2018)	Human activity recognition method based on inertial sensor and barometer,	The testing results, which accurately detect human activities based on the given sensor data, show how effective the system is.
10	E. Kim, S. Helal and D. Cook(2010)	Human Activity Recognition and Pattern Discovery	The study distinguishes between broad systems that recognize activity using predetermined models and those that use sensor data analysis.

2.2 Limitations in Existing Methods:

Current activity tracking tools still fall short in matching the fluid flexibility of human movement. Methods that work well in controlled settings often get tripped up by natural variety. The ultimate goal remains out of reach: systems that can keep up with the complexity of life.

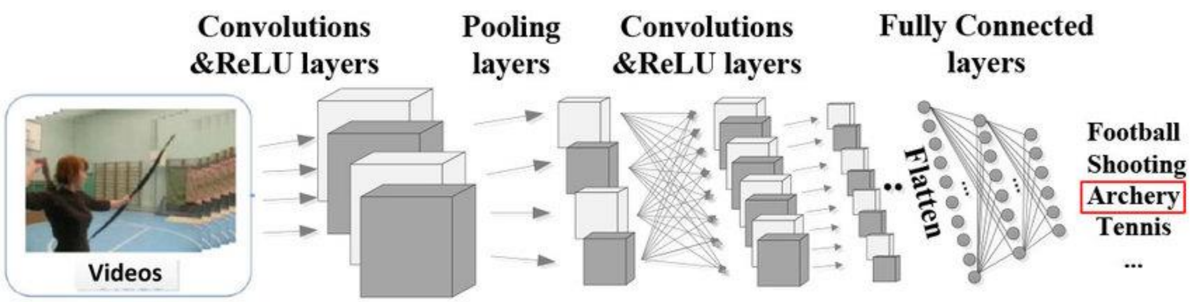
Real-time recognition still strains many algorithms, even though timely responses matter in safety applications. And we attachment trackers in so many unique places, but data quality and placement inconsistencies continue causing trouble.

There's still progress to make before activity tracking can adapt to diverse abilities, bodies, environments - all the richness of living. Closing these gaps will take creativity, but will uncover new potentials. More flexible, quick-thinking systems would revolutionize applications from healthcare to wearables and beyond. The open questions invite collaborative innovation so one day, trackers can handle life's every variation.

3. Proposed Method:

Firstly, the dataset is imported. The UCF101 dataset is used for the study. The video clips is to be loaded and authenticated for each of the 101 action categories. Form image patterns by taking the frames out of the video footage. This would be regarded for specific applications. Extract the features and choose appropriate CNN module for classification. The most suitable architecture for video classification is CNN I3D (Inflated 3D CNNs) and is hence used. The dataset used for training the data is UCF101. Hence, we check the labeling of the data with the use Kinetics and hence display it for ease access. The trained model is tested with the use of sample videos which then displays the probabilistic percentage of the possible labels associated with the video. The one with the highest probability is associated with the video and is hence verified. The testing of data is done using various video clips from the dataset.

4. System design



Firstly, the UCF101 dataset is downloaded and preprocessing is performed. This ensures that the data from the video is well segmented into required and desirable frames and clip length. Kinetics URL is used to authorize and label the data. Perform data preprocessing steps, such as resizing frames, normalizing pixel values, and potentially extracting optical flow or other relevant features. Split the dataset into training, validation, and test sets. 1CNN Model I3D is chosen for performing classification. This helps in reducing complexity and showcases rigid classification tasks within the video sequence. Here, we use UCF101 dataset and hence the model is already trained. However, we fine tune the dataset for adaptability by further initializing pre-trained weights. The use of Kinetics is made as it simplifies the task. The CNN model is then tested post training on the designated test data.

4.1: Results:

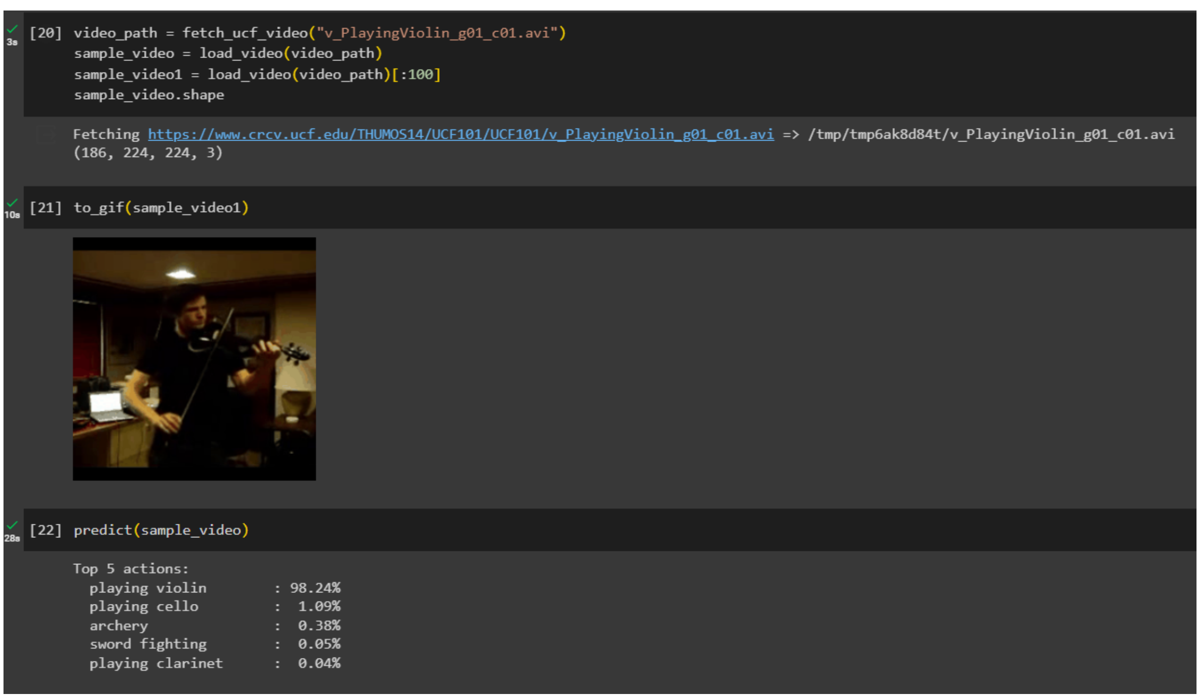
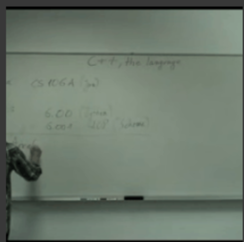


Fig 1.2: Testing the model for “Playing Violin” from UCF101 dataset. Result is satisfactory.

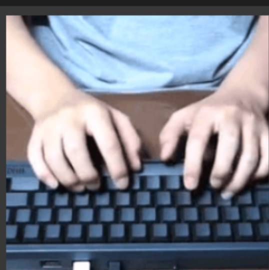
```
[29] video_path = fetch_ucf_video("v_WritingOnBoard_g01_c01.avi")
     sample_video = load_video(video_path)
     sample_video1 = load_video(video_path)[:100]
     sample_video.shape

     Fetching https://www.crcv.ucf.edu/THUMOS14/UCF101/UCF101/v_WritingOnBoard_g01_c01.avi => /tmp/tmp6ak8d84t/v_WritingOnBoard_g01_c01.avi
     (153, 224, 224, 3)

[30] to_gif(sample_video1)
```



```
     predict(sample_video)

     Top 5 actions:
       writing              : 99.85%
       spray painting       :  0.01%
       beatboxing           :  0.01%
       ice fishing          :  0.01%
       smoking              :  0.01%
```

Fig 1.2: Testing the model for “Writing on Board” from UCF101 dataset. Result is satisfactory.

```
[32] video_path = fetch_ucf_video("v_Typing_g01_c03.avi")
     sample_video = load_video(video_path)
     sample_video1 = load_video(video_path)[:100]
     sample_video.shape

     Fetching https://www.crcv.ucf.edu/THUMOS14/UCF101/UCF101/v_Typing_g01_c03.avi => /tmp/tmp6ak8d84t/v_Typing_g01_c03.avi
     (142, 224, 224, 3)

[33] to_gif(sample_video1)
```



```
[34] predict(sample_video)

     Top 5 actions:
       using computer         : 100.00%
       drumming fingers       :   0.00%
       texting                :   0.00%
       using remote controller (not gaming):  0.00%
       tapping pen            :   0.00%
```

Fig 1.3: Testing the model for “Typing” from UCF101 dataset. Result is satisfactory.

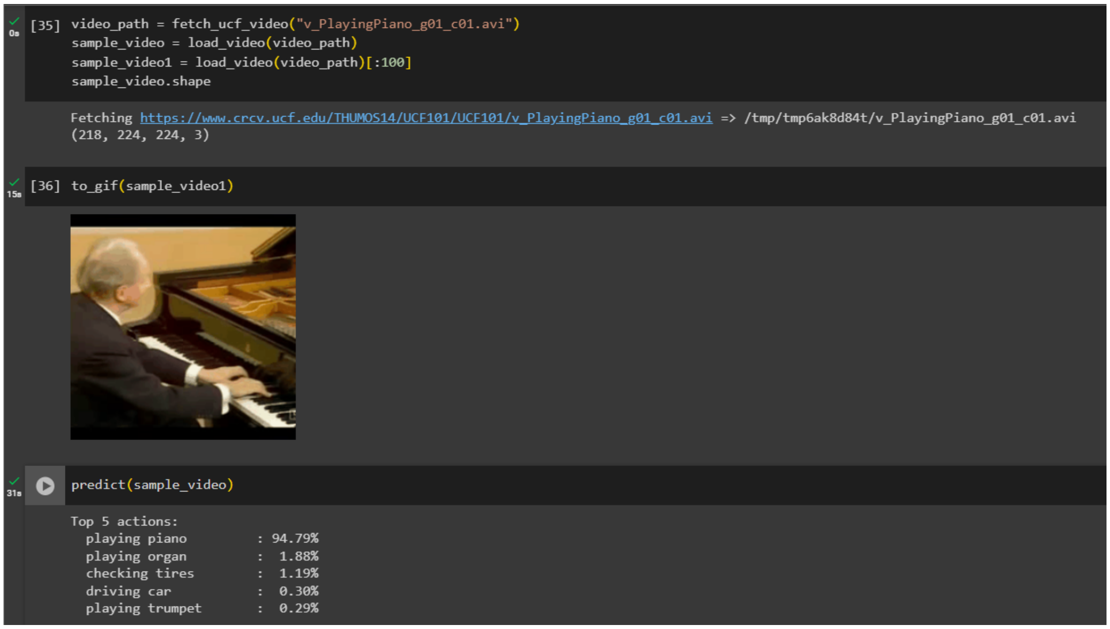


Fig 1.4: Testing the model for "Playing Piano" from UCF101 dataset. Result is satisfactory.

The above results show various Human Activities which are classified in the dataset and further labeled based on their likelihood of matching with the features extracted while training the model. The most suitable activity associated with video clip is labeled through Probability distribution. The activity with the highest probability is labeled to the video clip.

5. Conclusion:

In conclusion, the implementation of Convolutional Neural Networks (CNNs) into Human Activity Recognition (HAR) is an important advancement towards more accurate, versatile, and real-time activity recognition. The findings of this study highlight the potential of CNN architectures to boost recognition accuracy, encapsulate temporal dynamics, and minimize emphasis on handcrafted attributes. The proposed model is more resistant to environmental alterations and have a greater ability to generalize over a broad spectrum of human activities. Furthermore, research into multimodal sensor fusion and an emphasis on interpretability contribute to the overall progress of HAR with CNNs. As the model is adaptable, users and researchers can obtain useful insights into how the model makes its predictions, increasing transparency and confidence in the system's decision-making process, which is especially crucial for applications requiring human trust. The implementation of transfer learning enhances the ability of the model to adapt to the intricate nature of UCF101's activity classes by utilizing pre-trained weights. Post-processing techniques, such as temporal smoothing, help to refine predictions on video sequences, and the optional research of multimodal sensor fusion offers prospective pathways for increasing resilience. As the developed HAR system demonstrates its efficacy in recognizing actions in real-world video data, the results contribute not only to the field of activity recognition, but also to practical applications spanning security, healthcare, and human-computer interaction.

6. References:

[1] I. Stolovas, S. Suárez, D. Pereyra, F. De Izaguirre and V. Cabrera, "Human activity recognition using machine learning techniques in a low-resource embedded system," 2021 IEEE URUCON, Montevideo, Uruguay, 2021, pp. 263-267, doi: 10.1109/URUCON53396.2021.9647236.

[2] E. Kim, S. Helal and D. Cook, "Human Activity Recognition and Pattern Discovery," in IEEE Pervasive Computing, vol. 9, no. 1, pp. 48-53, Jan.-March 2010, doi: 10.1109/MPRV.2010.7.

[3] F. Zhou, R. Wang, H. Su and S. Xu, "A Human Activity Recognition Model Based on Wearable Sensor," 2022 9th International Conference on Digital Home (ICDH), Guangzhou, China, 2022, pp. 169-174, doi: 10.1109/ICDH57206.2022.00033.

[4] M. Atikuzzaman, T. R. Rahman, E. Wazed, M. P. Hossain and M. Z. Islam, "Human Activity Recognition System from Different Poses with CNN," 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2020, pp. 1-5, doi: 10.1109/STI50764.2020.9350508.

[5] L. Xie, J. Tian, G. Ding and Q. Zhao, "Human activity recognition method based on inertial sensor and barometer," 2018 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL), Lake Como, Italy, 2018, pp. 1-4, doi: 10.1109/ISISS.2018.8358140.

[6] F. Yang, X. Wen, Y. Geng, Y. Wang, X. Wang and C. Lu, "MPJA-HAD: A Multi-Position Joint Angle Dataset for Human Activity Recognition Using Wearable Sensors," 2022 International Conference on Advanced Mechatronic Systems (ICAMechS), Toyama, Japan, 2022, pp. 178-182, doi: 10.1109/ICAMechS57222.2022.10003441.

[7] Q. Jian, S. Guo, P. Chen, P. Wu and G. Cui, "A Robust Real-time Human Activity Recognition method Based on Attention-Augmented GRU," 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 2021, pp. 1-5, doi: 10.1109/RadarConf2147009.2021.9455322.

[8] L. Amin Choudhury, S. Moulik and S. Choudhury, "Cloud-based Real-time and Remote Human Activity Recognition System using Wearable Sensors," 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), Taoyuan, Taiwan, 2020, pp. 1-2, doi: 10.1109/ICCE-Taiwan49838.2020.9258050.

[9] Q. Saini and V. Maan, "Human Activity and Gesture Recognition: A Review," 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3), Lakshmangarh, India, 2020, pp. 1-2, doi: 10.1109/ICONC345789.2020.9117535.