1. Introduction

Human action recognition is an important and challenging task that has various real-world applications [1]. It can be used in video surveillance systems to automatically detect unusual or suspicious activities [2], in sports analysis to evaluate athletes' performance, and in healthcare to monitor patients' movements. Developing an accurate and efficient action recognition model significantly improves these applications' performance and reduce human intervention [3].

2. Related Work

Several studies have been conducted in the field of human action recognition, including hand-crafted feature-based methods [4] and deep learning-based approaches [3]. Two-stream convolutional neural networks (CNNs) [4] and 3D CNNs [5] have shown promising results in recognizing human actions from video sequences.

3. Dataset

The UCF101 dataset will be used for model building. It contains 13,320 video clips of human actions from 101 action categories [5]. The dataset has been extensively used in previous research on action recognition and is publicly available.

3.1. Sample size and manual labeling

The UCF101 dataset contains a sufficient number of video clips to build a robust action recognition model. The dataset has been previously used in several research studies that have achieved high accuracy in recognizing human actions [4]. No manual labeling is required, as the dataset is pre-labeled.

4. Method

The model aims to accurately classify human actions captured in video sequences, such as walking, running, jumping, and others. The model's output will be used to enhance video surveillance systems, sports analysis, and other applications that require real-time recognition of human actions.

4.1. High-level approach

I plan to build a 3D convolutional neural network (CNN) model to recognize human actions from video sequences. The model will take a sequence of frames as input and output the corresponding action class. I will pre-process the input data by resizing the frames, normalizing the pixel values, and applying data augmentation techniques to increase the dataset's size and variability. Then the model will be trained and evaluated using standard performance metrics, such as accuracy, precision, recall, and F1-score.

4.2. Peer reviewed article

"Two-Stream Convolutional Networks for Action Recognition in Videos" by [2] is a relevant peer-reviewed article to my project. The paper proposes a two-stream CNN architecture that captures spatial and temporal information separately from video sequences. Although my model's approach may differ from this paper, it provides a useful reference for designing a deep learning architecture for action recognition.

5. Timeline

Feb: Data pre-processing and augmentation
March1-15: Model development and training
March 15-30: Model evaluation, Report writing.

References

[1] Li, W., Li, X., Zhang, Y., Chen, X., & Yang, J. (2018). Action recognition based on deep learning: A review. Neurocomputing, 274, 410-425.

[2] Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Advances in neural information processing systems (pp. 568-576).

[3] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1725-1732).

[4] Soomro, K., Idrees, H., & Shah, M. (2012). Action recognition in realistic sports videos. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1-8).

[5] Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human action classes from videos in the wild. arXiv preprint arXiv:1212.0402.