A black background with white text

Description automatically generated with low confidence

MSc Data Science Project

7PAM2002-0509-2023

Department of Physics, Astronomy and Mathematics

**Data Science Project and Data Management Plan**

**Project Title:**

**Human Action Recognition using key point detection and MobileNetV2 Deep learning.**

**Student Name and SRN:**

Pavanteja Turaka and 21080899

Supervisor: Dr. William Cooper

**Aim of the project**

The objective of this project is to create a resilient Human Action Recognition (HAR) system by using key point detection based on Alpha Pose model and the MobileNetV2 deep learning architecture. This system will precisely recognize and categorize human motions from images by using alpha pose to monitor and analyse human body movements and MobileNetV2 for efficient and effective extraction and classification of features. The project aims to attain a high level of precision in recognizing actions in real-time, while also ensuring computing efficiency.

**Research questions:**

* How can key point detection accuracy be optimized when integrated with MobileNetV2 for human action recognition?
* How does the integration of temporal information from key point sequences improve the action recognition capabilities of a MobileNetV2-based model?
* What are the performance trade-offs between accuracy and computational efficiency when using MobileNetV2 for key point-based human action recognition?

**Objectives of project:**

Perform an extensive examination of existing literature:

* To conduct a comprehensive examination of the current body of literature on Human Action Recognition (HAR) approaches, the primary focus will be on the techniques used for key point identification and the use of deep learning models, with specific emphasis on MobileNetV2. The process will include examining recent progress, identifying deficiencies in existing research, and providing the theoretical basis for the planned investigation.

Create a framework for detecting key points.

* The objective is to create and execute a strong framework that can identify crucial moments in human body motions by using cutting-edge critical point recognition algorithms. This framework will function as the primary phase of the HAR system, recording exact movement patterns that are crucial for precise action identification.

Implement the MobileNetV2 model for the purpose of recognizing actions.

* The objective is to include the MobileNetV2 deep learning architecture into the key point detection framework to achieve efficient and precise categorization of human activities. This aim involves enhancing the efficiency of the MobileNetV2 model for analysing key point data and assessing its performance in comparison to other deep learning models.

Analyse system performance and enhance optimization:

* To methodically examine the performance of the integrated Human Activity Recognition (HAR) system in different real-world situations, by evaluating parameters such as accuracy, latency, and computing efficiency. This will need doing experiments using various datasets, using data augmentation approaches, and investigating possible advancements via hyperparameter tweaking and model upgrades.

Record and present the discovered information and conclusions:

* To provide a thorough report that provides a detailed explanation of the study methodology, experimental setup, findings, and conclusions. The report will include a meticulous evaluation of the results, deliberations on the ramifications of the study, and suggestions for further research. In addition, it will compile the literature research, design procedures, and performance assessments, creating a comprehensive academic reference for the built HAR system.

**Background**

Human Action Recognition (HAR) seeks to automatically analyse and identify the characteristics of an action from unfamiliar video sequences. The increasing need for automated interpretation of human behaviour has led to a significant interest in Human Activity Recognition (HAR) in both academic and industrial sectors. Indeed, the analysis and comprehension of an individual's behaviour is essential for a diverse array of applications, including video indexing, biometrics, surveillance, and security.

Zhu, Q., et al,(2018) suggested a semi-supervised deep learning method for implementing temporal assembling of deep long-short term memory (DLSTM) on both labelled and unlabelled data. The data and properties of smartphone inertial sensors were collected using a deep neural network (DNN) to analyse local relationships. The researchers analysed their findings with many algorithms assessed on the UCI dataset to provide cutting-edge outcomes.

Chelli, A. and Pätzold, M., (2019) created a system capable of identifying seven distinct actions, one of which is the detection of falls. Acceleration and angular velocity data were obtained from a cell phone, and time and frequency domain characteristics were extracted from this data. The system demonstrated an accuracy of 81.21% using Artificial Neural Networks (ANN), 87.8% with K-nearest neighbour (KNN), 93.21% with Quadratic Support Vector Machines (QSVM), and 94.12% accuracy using ensemble bagged tree (EBT) while analysing simply acceleration data. The inclusion of characteristics derived from acceleration and angular velocity significantly enhances the accuracy of the algorithms, resulting in improvements of 85.8%, 91.8%, 96.1%, and 97.7%. Notably, the fall detection accuracy of QSVM and EBT achieves a perfect 100% without any false alarms, representing the highest achievable performance.

Mohan, A., et al (2019) introduced a technique for manually observing and tracking uncommon abnormal behaviours in supermarkets, public spaces, and university campuses. Principal Component Analysis (PCA) and Convolutional Neural Networks (CNN) address the issue of manual methods, such as false alarms, by accurately identifying and pinpointing irregularities in video footage. PCA and SVM classifier are used to identify abnormal occurrences on a frame-by-frame basis. The suggested technique achieved superior outcomes on the UCSD, UMN dataset, and Avenue Dataset, surpassing all previous methods. Shoplifters may effortlessly detach labels from merchandise even while under the surveillance of Electronic Article Surveillance (EAS) devices. The CNN model receives live video feed from CCTV cameras to identify potentially illicit human behaviours inside the business premises, including stealing, robbery, and break-ins, and subsequently triggers an alert. The suggested approach surpasses others with a precision rate of 89%.

Xia, L. and Li, Z., (2021) used a complete Convolutional Neural Network (CNN), specifically a pre-trained VGG-16 model, to extract static appearance characteristics. The temporal attention mechanism captures appearance characteristics at same positions. The LSTM network used these attributes to infer aberrant features and detect abnormal activity in video frames. The suggested technique outperformed existing methods in terms of pixel and frame-level outcomes.

**Data Management Plan**

**About dataset:**

Human Action Recognition (HAR) seeks to comprehend human behaviour and provide a categorical designation to each action. Due to its diverse variety of applications, computer vision has been garnering growing interest in the industry. The collection contains 15 distinct categories of Human Activities. The dataset consists of around 12,000 labelled photos, which includes the validation images. Each photograph is assigned to a single human activity category and is stored in a separate folder according to its labelled class.

The dataset contains images of various people doing 15 different activities, like sleeping, talking, etc. The shape and size of the image are different from each other, necessitating reshaping for the algorithm. The dataset in total has a size of 350 MB, and each image is between 10 kb and 25 kb, depending on its quality.

I have chosen this HAR dataset due to the following reasons: HAR dataset include activities recorded in real-world settings, which aids in the development of resilient models that excel in actual use cases. HAR datasets provide as a standardized baseline for evaluating and comparing various algorithms and models, hence promoting innovation and advancements in the area. Datasets including annotations for real-time situations are valuable for the development and testing of human activity recognition (HAR) systems that can function efficiently in live applications, such as surveillance and human-computer interaction.

The data is downloaded from Kaggle and is publicly available for download and using in research projects.

Link: <https://www.kaggle.com/datasets/meetnagadia/human-action-recognition-har-dataset/data>

A person riding a bicycle

Description automatically generated A person holding a cell phone

Description automatically generated

Figure 1: Sample images.

**Document control:**

GitHub Link:

I will be using the above GitHub link for updating and version-controlling the documents. The repository will contain all the documents and code related to the project. The updating frequency will be once a week; the whole code and documentation will be updated, and the records will also be logged in Herts logbook. I will store the second copy of the code and documentation on my personal storage drive. I will also add my supervisor as a viewer to the GitHub repository, enabling him to monitor the progress of the work. Additionally, I will maintain the repository's public access for other collaborators.

The readme file will contain important information about who the code is written for, the libraries it requires for execution, and other requirements.

**Ethical requirements**

* The data complies with GDPR standards, ensuring that all personal data is handled in accordance with the EU's stringent data protection regulations.
* The project adheres to the University of Hertfordshire's ethical guidelines, guaranteeing that all research activities are conducted with integrity and respect for ethical principles.
* We have obtained permission to utilize the data for my intended research purposes, ensuring that our use of the data is authorized and appropriate.
* There is a strong assurance that the data was gathered ethically by the original collectors, ensuring that the methods of data collection, collation, and creation met ethical standards.

**Project plan**

**Ghant Chart**

Table 1: Ghant Chart

A blue and white calendar with a graph

Description automatically generated with medium confidence

The project is divided into 8 tasks, which are planned to be achieved in a span of 3 months. These tasks aim to achieve the objectives of the project and keep track of it.

**PDM submission**: this task is associated with the submission of the 1st draft of the PDM version in the 1st week of July and the submission of the final document before the 17th of June. This submission of the report is a milestone for the task. Ethics quiz: This task is associated with preparing and attending an ethics quiz. I have allocated around 4 to 5 days.

**Literature review**: This task is associated with reading papers on human action recognition and how key points and deep neural network methods were developed. This includes defining a search string with inclusion and exclusion conditions. The key milestone of this task is to find gaps in the literature and act on them.

**Data collection**: download the dataset from the link.

**EDA of the dataset**: This task is associated with data analysis of image data, like performing histogram analysis of intensity, checking for balance in the in the dataset, and many more. The milestone of this task is to find out more about the dataset and decide what preprocessing it requires.

**Data preprocessing**. This task is associated with improving the image quality and enhancing it. By using the inputs from the EDA task, perform appropriate preprocessing of the dataset. For example, balancing the dataset, reducing brightness, augmentation, etc. The milestone of this task is to find the best preprocessing method for images.

**Model development**: this task is associated with the development of the proposed model and training the model with the pre-processed dataset. The model will be trained and tested on various parameters and metrics. The milestone of this task is to get a working model with good accuracy.

**Model testing**: the model will be tested on various conditions, like without the use of key point detection, other data preprocessing techniques, using different weights for the model, etc. All the models will be compared with each other, and the best model will be selected.

**Report writing**: This task is associated with the documentation of all the findings of the project.

**Reference**

Xia, L. and Li, Z., 2021. A new method of abnormal behaviour detection using LSTM network with temporal attention mechanism. The Journal of Supercomputing, 77(4), pp.3223-3241.

Mohan, A., Choksi, M. and Zaveri, M.A., 2019, July. Anomaly and activity recognition using machine learning approach for video-based surveillance. In 2019 10th international conference on computing, communication and networking technologies (ICCCNT) (pp. 1-6). IEEE.

Chelli, A. and Pätzold, M., 2019. A machine learning approach for fall detection and daily living activity recognition. IEEE Access, 7, pp.38670-38687.

Zhu, Q., Chen, Z. and Soh, Y.C., 2018. A novel semi supervised deep learning m0ethod for human activity recognition. IEEE Transactions on Industrial Informatics, 15(7), pp.3821-3830.