FYP PROGRESS REPORT

Skateboard Trick Recognition through an AI-based Approach

Kris Saliba

Supervisor: Dr. Joseph Bonello

January 2024

Submitted in partial fulfilment of the requirements for the degree of Bachelor of Science in Information Technology (Honours) (Software Development).

L-Università ta' Malta
Faculty of Information &
Communication Technology

Contents

List of Figures

1 Introduction

1.1 Introduction to Skateboarding and its Evolution

Skateboarding dates back to the 1940s when handmade skateboards first appeared [1]. It has since developed into a worldwide phenomenon, with its popularity skyrocketing, after gaining recognition as an official sport in the 2020 Tokyo Olympic Games [2]. Skateboarding comprises the dynamic activities of riding a skateboard and skillfully performing a repertoire of tricks, manifesting as a popular and exhilarating “extreme sport”.

This dynamic sport encompasses various disciplines and riding styles, each offering unique challenges for skateboarders to explore. Two of the most prominent styles are “vert” and “street.” Vert skateboarding revolves around riding on specialised obstacles, namely, half-pipes and ramps, emphasising aerial manoeuvres. Street skateboarding transpires in urban environments, utilising various obstacles that can be found outdoors, including stairs, rails, ledges, gaps or flat ground for skaters to showcase their creativity [3].

1.2 Motivation

The current method of identifying tricks in skateboard competitions relies heavily on judges verbally stating them during live broadcasts. This approach lacks a digital overlay that could display the performed trick for the viewers, leading to a reliance on subjective judgment. Such subjectivity can lead to scoring disparities and conflicts, undermining competition fairness. Furthermore, this absence of an objective identification method obstructs the skaters’ ability to receive real-time feedback, which could be beneficial for skill development. An AI-based system that delivers a dependable and efficient method of trick recognition could improve this aspect and contribute to the lack of study in this area.

2 Scope

The scope of this study is not only to develop a skateboard trick classifier but also to delve into the largely unexplored field of Artificial Intelligence (AI) in action sports. Studying AI in this domain opens up possibilities for innovation in the skateboarding industry, such as introducing fairer scoring systems in competitions and enhancing the live-streaming viewing experience.

Moreover, this study offers various research opportunities, like providing detailed coaching feedback, where, in the case of skateboarding, Machine Learning (ML) could detect inaccuracies in body weight placement and foot positioning. Additionally, this study opens avenues for injury prevention research by identifying common injury patterns and risky manoeuvres.

3 Brief overview of the literature

Recent advancements in recognising human actions in videos have significantly impacted various fields, ranging from the medical sector to surveillance systems, [4], [5]. Particularly noteworthy is its application in developing an AI-based skateboard trick classifier, an area that has seen limited research.

In this emergent field, leveraging activity recognition techniques from a video has led to two primary methodologies among researchers. The first technique involves utilising signals obtained from skateboard-mounted accelerometers or signals artificially generated based on the findings of prior studies. These signals are then fed into a study-dependent model for classification, as outlined in [6] and [7]. The second approach employs computer vision techniques, leveraging video footage of skateboard tricks to train and refine models for accurate trick identification, as depicted by the studies [8] and [9].

3.1 Accelerometry approach

The study by Abdullah et al (2021) [6], makes use of a custom dataset comprising of six skateboard tricks most commonly executed in competitive events. Amateur skateboarders performed each trick five times on a modified skateboard equipped with an Inertial Measurement Unit (IMU) to record the signals produced. The researchers capture six signals for each trick, including linear accelerations along the x, y, and z axes (aX, aY, aZ) and angular accelerations along the same axes (gX, gY, gZ). They then opt for the unique approach of concatenating all six signals onto a single image corresponding to one trick, employing two input image transformations: raw data (RAW) and Continuous Wavelet Transform (CWT). With the application of six transfer learning models on this data, the results show that RAW and CWT input images on MobileNet, MobileNetV2 and ResNet101 models achieved favourable accuracy. However, the CWT-MobileNet-Optimised SVM pipeline was deemed the best due to its reduction in computational time.

On the other hand, the study by Corrêa et al (2017) [7], obtained their sample data by artificially generating 543 signals based on prior research, utilising tools such as MATLAB 2015 and Signal Processing Toolbox. These signals were then categorised into five distinct classes representing different skateboard tricks, each with various samples ranging from 30 to 50 per class, across three axes (X, Y and Z). This study developed and validated individual Artificial Neural Networks (ANNs) for each axis, as well as the combination of the three: ANN XYZ, displaying the potential of Neural Networks to categorise multidimensional skateboard tricks. The ANNs are all multilayer feed-forward neural networks (MFFNNs), structured into three distinct layers. They feature an input layer with 82 neurons, a hidden layer, comprised of 23 neurons utilising a tan-sigmoid transfer function and an output layer consisting of 5 neurons with a softmax function. Finally, the study achieved high accuracies, with ANNs X, Y and Z achieving accuracies of 94.8%, 96.7% and 98.7%, respectively, while the combined ANN XYZ achieved an accuracy of 92.8%.

3.2 Computer Vision approach

The paper by Shapiee et al (2020) [8] leverages a custom data set comprising videos capturing the execution of five distinct skateboard tricks, each attempted five times. Each video spans two to three seconds, yielding a total of 750 images by extracting 30 frames per video. This study made use of data augmentation techniques to expand their dataset further. Consequently, they introduced an additional 2,250 images, achieving 3,000 images in their data set. On the other hand, Chen (2023) [9] compiled a comprehensive data set by collecting videos from multiple platforms, including YouTube, Twitter and Instagram. Furthermore, Chen trained the model using 15 fundamental tricks commonly observed in competitive settings. The researcher collected 50 videos per trick, summing up to a total number of 750 videos. Of these, 45 videos per trick were allocated for training, and the remaining 5 were reserved for validation.

Data augmentation techniques are popular in studies with relatively limited datasets. Techniques such as flipping, rotating, scaling and colour manipulation not only enhance the size of the original dataset but also lower the likelihood of the model overfitting [10]. The paper by Shapiee et al [8] utilises three rotation augmentation techniques: horizontal rotation, positive 90°rotation and negative 90°rotation. The researchers experimented on three Transfer learning models: MobileNet, NASNetMobile and NASNetLarge, each evaluated using a k-Nearest Neighbor (k-NN) classifier. As a result, the models demonstrated impressive classification accuracies, with MobileNet achieving 95%, NASNetMobile 92% and NASNetLarge 90%.

Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs) are popular architectures due to their capabilities in modelling the dynamic relationships in sequential data [11]. In the student abstract by Hanciao Chen [9], extensive experimentation is conducted using diverse models, exploring various combinations of CNN-LSTM and CNN-BiLSTM, including integrating attention and transfer learning approaches. This study further documents and analyses important metrics such as training time, training accuracy and validation accuracy for each model experimented on. Among these, the top three models that stood out in terms of validation accuracy were the ResNet50 with Attention and BiLSTM (84%), ResNet50 with BiLSTM (81%) and ResNet50 with LSTM (80%). Chen's study provides valuable insight into the application of diverse models in activity recognition in skateboarding.

4 Aims and Objectives

The aims and objectives of this project are to design and develop a dependable Artificial Intelligence (AI) model that can identify various skateboard tricks directly from videos. An initial focus will be placed on the recognition of three distinct tricks namely, kickflips, ollies and shuvits. To reach this goal the appropriate ML techniques must be utilised to overcome basic computer vision challenges, such as varying camera angles, lighting conditions and the fast-paced nature of skateboard tricks. By addressing these challenges, this project aims to establish a usable tool that can be used by skateboarders, coaches, and skateboard competitions.

4.1 List of deliverables

- Extensive dataset compilation
- Trained model with detailed accuracy metrics
- Documentation of experiments and their visualisations

5 Methodology and Evaluation Strategy

In terms of data collection, this project will utilise a combination of the open-source datasets provided by Fitzgerald et al (2020) [12] and Chen [9]. Moreover, to ensure an unbiased and rigorous evaluation of the model's performance, this study will incorporate a dataset, consisting of skateboard tricks performed and recorded specifically for this study (by myself).

During my thesis research, I initially began by experimenting with various preprocessing techniques, such as applying frame extraction to videos and employing object detection to selectively crop each frame, focusing on encapsulating only the skater. This approach was designed to remove any unnecessary overhead in each frame that might otherwise impede the learning process.

I commenced the development phase using Python, opting for YOLO (You Only Look Once) as the object detection framework. The YOLO training process involved gathering a diverse dataset and labelling each image featuring a skateboarder by drawing a bounding box around each subject. The trained object detection model takes an image as an input and returns the coordinates of the bounding box encasing the skater's position, used to crop the image accordingly.

After completing the experimentation phase with various preprocessing techniques and creating a temporary preprocessing mechanism, I developed a baseline model. A baseline model establishes a benchmark for all future modifications and enhancements to it. This foundational model, characterised by its ConvLSTM architecture, is designed explicitly for sequential data processing. It begins with a ConvLSTM2D input layer, which features a 3x3 kernel size and employs a 'tanh' activation function. Subsequently, this model incorporates a MaxPooling3D layer with a pool size 1x2x2, complemented by TimeDistributed Dropout layers to prevent overfitting. The architecture consistently follows this pattern, progressively increasing the number of filters, finally ending with a Dense layer. After training, the baseline model achieved a favourable but improvable validation accuracy of 72%.

5.1 Work Plan

I plan to refine the baseline model further and aim to achieve a more efficient and accurate model. The approach will involve applying various preprocessing and training techniques, including but not limited to data augmentation and the 'leave-one-out' approach. Additionally, I intend to experiment with transfer learning methods, incorporating models like ResNet or VGG for feature extraction and exploring adjustments in architectures and hyperparameters.

5.2 Project timeline

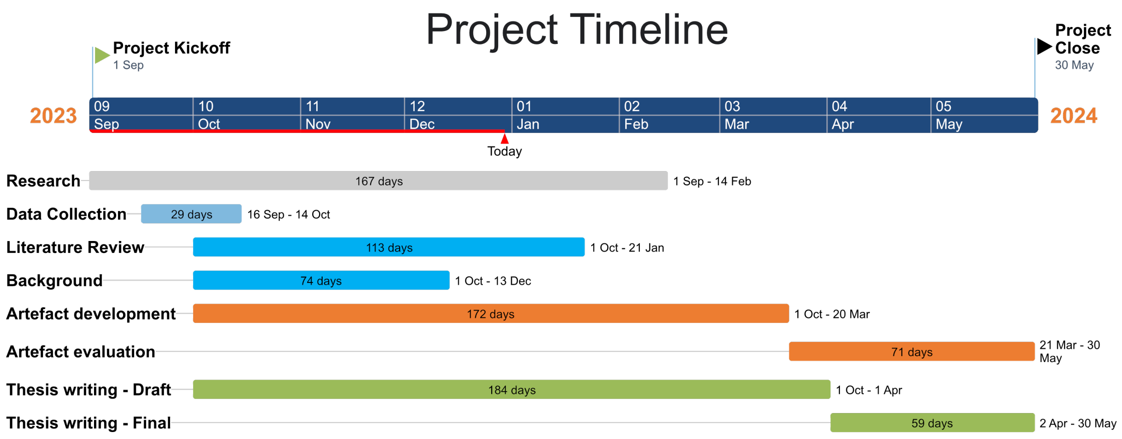

Figure 5.1 Estimated project timeline.

Bibliography

[1] *Encyclopaedia britannica*, 2023. [Online]. Available: https://www.britannica.com/sports/skateboarding.

[2] *One year on: How skateboarding's olympic debut changed the games 2022*, 2022. [Online]. Available: https://olympics.com/en/news/one-year-on-skateboarding-olympic-debut-feature.

[3] Z. Foley, *Vert, street, park - what are the different styles of skateboarding?* 2021. [Online]. Available: https://www.goskate.com/top/skateboarding-styles-full-guide/.

[4] J. Yin *et al.*, "Mc-lstm: Real-time 3d human action detection system for intelligent healthcare applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 2, pp. 259–269, 2021. DOI: 10.1109/TBCAS.2021.3064841.

[5] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008. DOI: 10.1109/TKDE.2007.1042.

[6] M. A. Abdullah *et al.*, "The classification of skateboarding tricks via transfer learning pipelines," *PeerJ Computer Science*, vol. 7, e680, 2021.

[7] N. K. Corrêa, J. C. M. d. Lima, T. Russomano, and M. A. d. Santos, "Development of a skateboarding trick classifier using accelerometry and machine learning," *Research on Biomedical Engineering*, vol. 33, pp. 362–369, 2017.

[8] M. Shapiee *et al.*, "The classification of skateboarding tricks: A transfer learning and machine learning approach," *MEKATRONIKA*, vol. 2, pp. 1–12, Oct. 2020. DOI: 10.15282/mekatronika.v2i2.6683.

[9] H. Chen, "Skateboardai: The coolest video action recognition for skateboarding (student abstract)," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 16 184–16 185, Jun. 2023. DOI: 10.1609/aaai.v37i13.26952.

[10] S.-A. Rebuffi, S. Gowal, D. A. Calian, F. Stimberg, O. Wiles, and T. A. Mann, "Data augmentation can improve robustness," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, Curran Associates, Inc., 2021, pp. 29 935–29 948. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/fb4c48608ce8825b558ccf07169a3421-Paper.pdf.

[11] R. C. Staudemeyer and E. R. Morris, *Understanding lstm – a tutorial into long short-term memory recurrent neural networks*, 2019. arXiv: 1909.09586 [cs.NE].

[12] LightningDrop and C. Fitzgerald, *Skateboardml 1.0*, version 1.0, Aug. 2020. DOI: 10.5281/zenodo.3986905. [Online]. Available: https://doi.org/10.5281/zenodo.3986905.