Swing and Head Synchronisation in Golf using Video Analysis
Literature Review

MARC MARAIS
Supervisor: DR. DANE BROWN
Computer Science Department, Rhodes University

May 21, 2021

Abstract

Computer vision in golf is commonly found using ball-tracking technology, but what about extending this to player swing and head posture sequencing? Thus, enabling the amateur golfer to pick up when their swing and head position are out of sync and improve this aspect of their game. Using an existing dataset GolfDB and a combination of image processing and machine learning techniques, an attempt will be made to achieve this. Action recognition is the core aspect of making this possible. Various image processing techniques may be needed to extract the golf club from the video frames. Determining the best-fit machine or deep learning model will be vital. From this, the model will be trained and tested to improve prediction accuracy when checking for synchronisation between head position and swing sequence.

1 Introduction

Golf is one of the more complex and challenging sports around today and is highly popular as more than 80 million people play the sport worldwide (McNally et al., 2019). The fine-tuning and complexity behind the golf swing itself comes under much scrutiny. With little research or studies on the effect of different swings, much of the debate regarding the correct body posture and swing technique stems from professional golfers and coaches personal opinions (Smith et al., 2012). Correct head and golf ball alignment is one of the key elements to a good swing (Glazier and Lamp, 2013).

The golf swing action can be recognised using machine learning, deep learning, and neural network algorithms. This study will attempt to pick up correct and incorrect head alignment during a golf swing. Golf swing sequencing has been achieved using a lightweight deep neural network (McNally et al., 2019). Using this study, which uses a labelled video database to sequence golf swing events, head alignment will be checked for synchronisation with the swing sequence events.

According to Glazier and Lamp (2013) head movement during the golf swing is linked to body movement. It is virtually impossible to keep the head completely still during the golf swing but avoiding excessive head movement is necessary.

Many studies around golf swing sequencing use sensors and optical cameras to detect the swing pattern and events (Ko and Pan, 2020; McNally et al., 2019). Sensors and optical cameras with hardware processing are costly implementations to study swing events and sequencing. On the other hand, the use of computer vision and deep learning to analyse golf swing events is more practical and accessible to amateur golfers.

Therefore, a cost-effective and easily accessible solution that lies in the field of computer vision is necessary to aid golfers in the analysis of their swing and head movement.

1.1 Research Statement

A labelled dataset with a supervised machine learning model can be used to analyse swing and head synchronisation in golf in videos. A system, based on this model, can be constructed to assist both amateur and professional golfers in preventing incorrect head position during their swing by detecting these anomalies. High accuracy in anomaly detection can be compared to related system to measure the systems effectiveness.

2 Literature Review

Computer vision assisted analysis in sports is growing and used by broadcast viewers and in training and coaching. Tracking players and ball tracking forms a large part of computer vision (Thomas et al., 2017). In the golfing world, the use of ball-tracking to track a shot's trajectory has come in handy in assisting viewers in following the shot trajectory.

2.1 Image Processing

Image processing has been developed to implement features such as picture enhancement and restoration as well as picture segmentation which can be linked to machine learning (Petrou and Petrou, 2010).

Gehrig et al. (2003) presented an image processing technique to extract a golf club from a video frame.

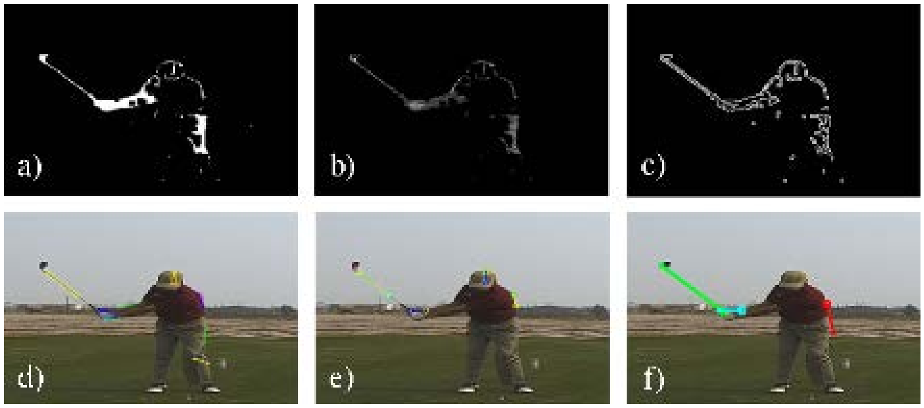

Figure 1: Golf club image extraction technique

To extract the the golf club swinging motion from the video frame, a binary mask is applied over the motion detection, as can be seen in Figure 1a (Gehrig et al., 2003). The next step is to greyscale the frame. Using the Canny method of edge detection, extraction of the straight aspects of the image such as the golf club shaft and arms can take place (Canny, 1986; Gehrig et al., 2003). From the extracted edges, close parallel segments are merged into one component.

The success rate of this study was promising, although complications did occur. As a typical golf swing is a fast action and a club shaft is thin, a high-velocity swing may further blur the club shaft (Gehrig et al., 2003).

2.2 Machine Learning Network Architecture

Machine Learning is a vast field of computer science focused on algorithms and techniques designed to train and make predictions based on large datasets (Lee, 2019).

2.2.1 Machine learning in Golf

Research conducted, in the field of golf swing analysis, by McNally et al. (2019) aimed to create a sequence of golf swing events through localising each event to a single frame. Events included the address, toe-up, mid-backswing, top, middownswing, impact, mid-follow through and finish. These events were analysed and enabled consistent evaluation of the golf swing. The image processing methods, applied to the ImageNet dataset (Deng et al., 2009), to extract these events frame-by-frame were implemented using techniques such as bilinear interpolation[1] and normalisation[2] by subtracting the ImageNet means and dividing by the ImageNet standard deviation.

Using a benchmark database GolfDB, which consists of 1400 labelled videos and a combination of SwingNet, McNally et al. (2019) implemented a lightweight deep learning neural network to analyse the performance of golf swings. SwingNet compromises a network architecture design, incorporating lightweight convolutional neural networks (CNNs) to enable effective mobile deployment (McNally et al., 2019).

CNNs form part of the artificial neural network (ANN) architecture. CNNs focus area is to assist with solving complex image pattern recognition tasks (O'Shea and Nash, 2015). CNNs are very effective when implemented with labelled datasets and combined with feature extraction systems such as SIFT (scale-invariant feature transform) or HoG (Histogram of Oriented Gradients)[3] (LeCun et al., 2010).

O'Shea and Nash (2015) differentiates ANNs from CNNs in that CNNs use neurons that self-optimise through learning. The only drawback with CNNs is the large datasets required to train the model, which opens the door to overfitting. Overfitting reduces the model's ability to generalise and predict effectively in training, test, and prediction data sets (O'Shea and Nash, 2015).

A comparison was drawn regarding the accuracy of real-time vs slow-motion videos. Real-time videos were sample at 30 frames per second (fps), whereas the native frame rate of slow-motion was unknown. A scaled sample-dependent tolerance of the frames between the address and impact stage of the swing

[1]Understanding bilinear interpolation further reading at `https://www.cambridgeincolour.com/tutorials/image-interpolation.html`

[2]Normalisation involves changing the range of the pixel intensity of an image (Gonzalez, 2008)

[3]HoG is used to achieve object detection by counting occurrences of gradient orientation in localized portions of an image (Dalal and Triggs, 2005)

approximated fps at 30 (McNally et al., 2019). The detection of the golf swing events had a higher detection rate using real-time videos as the Percentage of Correct Events (PCE) for real-time was 79.2% compared to 72.5% PCE for slow-motion videos (McNally et al., 2019). Overall, the study found that the CNN SwingNet averaged a rate of 76.1% (average of real-time and slow-motion PCE) at detecting all eight events in the golf swing and 91.8% at picking up six of the eight events. These impressive results provide a broad benchmark and a foundation for predicting human posture and synchronising head movement with golf swing events.

Another study conducted by Ko and Pan (2020) also looked into swing sequencing but included body sway analysis during the swing. Ko and Pan (2020) made use of a single frontal facing camera as well as a motion capture to perform 3D analysis of the swing and body motion.

Classification of the swing events and quantitative information on twisting angles of the head, upper body, pelvis and shoulder were implemented through a deep learning regression model (Ko and Pan, 2020). The regression model is based on a bi-directional long short term memory (Bi-LSTM) neural network [4]. In the forward direction, sequential temporal features are implemented. In the backward direction, temporal features are incorporated. Both forward and backward features improve classification performance.

Ko and Pan (2020) used a concept of golf known as head-up analysis, the changing of the central axis height and raising the head before impact with the golf ball. This common mistake made by amateur golfers causes the golfer's swing to become unbalanced, leading to a poor shot in either the form of an offset trajectory or a missed shot (Cooper et al., 1974).

Swing events are captured and extracted as a sequence of images using the Sequence Feature Extraction and Classification Network (SFEC-net) (Ko and Pan, 2020). SFEC-net is made up of three convolution layers, three pooling layers, and two fully connected layers. After extracting the swing event im-

[4] Bi-LSTM models fall into the category of Bidirectional Recurrent Neural Networks BRNN developed by Schuster and Paliwal (1997)

ages, head-up analysis is conducted, creating three-dimensional and three-axis rotation angles of head movement.

Table 1: Results for golf swing event classification using SFEC-net (Ko and Pan, 2020)

Section	**Accuracy**	**Recall**	**Precision**
1	93.92	0.94	0.98
2	96.92	0.97	0.93
3	92.16	0.92	0.89
4	89.24	0.89	0.95
5	84.88	0.85	0.76
6	97.01	0.97	0.95
7	99.01	0.99	0.99
Total	95.44	0.93	0.92

The results in Table 1 show that the SFEC-net accurately predicts the swing events relatively well achieving an average accuracy of 95.44% and precision of 92.21%. The use of many hidden units may increase the chance of overfitting, and therefore the need to tune hyperparameters exists to avoid overfitting (Cui and Bai, 2019). This study performed well, but a way to remove the need for the motion capture suit needs to be found to make the technology readily available to all golfers.

Both the studies conducted by McNally et al. (2019) and Ko and Pan (2020) use different techniques to achieve swing sequencing. The advantage of McNally et al. (2019)'s system is the access to the large dataset, GolfDB, to use when training and testing the model.

2.2.2 State-of-the-art Machine Learning

State-of-the-art machines tend to take a deep learning approach. Implying that the model handles all the work, thus alleviating the need for image processing (Schmidhuber, 2015).

Shinde et al.'s proposal to use YOLO (You Only Look Once) combined with LIRIS Human Activities dataset Wolf et al. (2014) for fast human action recognition and localisation is an alternative to general CNNs (Shinde et al., 2018).

Object detection in YOLO uses a single CNN instead of the general two-stream CNN, and thus requires fewer frames to achieve action recognition and localisation(Shinde et al., 2018). Shinde et al.'s version of YOLOv2 (Redmon and Farhadi, 2017) architecture comprises 24 convolutional layers and two fully connected layers.

The LIRIS dataset (Wolf et al., 2014) contains action categories relating to Human-Human interactions, human-object interactions and human-human-object interactions (Shinde et al., 2018).

The results obtained from Shinde et al.'s research indicate all three indicator values, precision, recall and F1-score, to achieve a good classification model are above 80%. When contrasted to other state-of-art methods such as Mukherjee et al. (2015)'s hybrid motion optical flow combined with a binary SVM for classification, which used the same LIRIS dataset, YOLO's F1-score of 88.358% outperformed Mukherjee et al.'s F1-score of 81.27%. Therefore, YOLO achieved better performance while using fewer frames to do so.

Building on previous work relating to classifying object proposals using deep convolutional networks Girshick proposes the use of Fast Region-based Convolutional Networks for object detection (Fast R-CNN).

The reasoning behind developing Fast R-CNN being that current methods using R-CNN Girshick et al. (2014) and SPPnet He et al. (2015) have drawbacks relating to training in a multi-stage pipeline, expensive use of space and time during training and slow object detection.

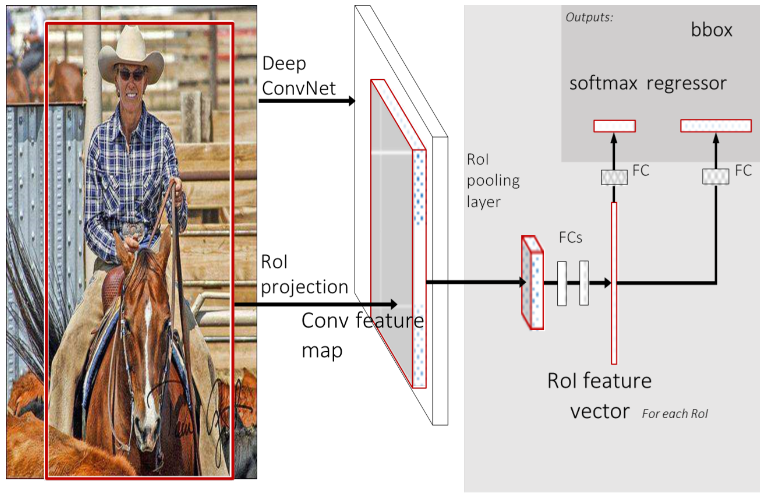

Figure 2: Fast R-CNN architecture (Girshick, 2015)

Fast R-CNNs architecture, as shown in figure 2 inputs an entire image and a set of object proposals. Several convolutional and max-pooling layers process the image to produce a convolutional feature map. A region of interest (RoI) pooling layer from feature map fixed-length feature vectors is extracted using the object proposals. The feature vectors are then split up into two classes of outputs softmax probabilities estimates over K object classes and a layer of four real-valued numbers for each K object class. These values compromise the bounding box positions.

R-CNN is advantages due to higher detection quality, single-stage training (using multi-task loss), training that can update all network layers and feature caching does not require disk storage Girshick (2015).

Based on the impact of the literature, investigation into state-of-the-art object detection models such as YOLO and R-CNN remains prudent (Girshick, 2015).

2.2.3 Outside the scope of Golf

Moodley and van der Haar (2020) proposed a cricket stroke recognition model to differentiate between the different classes of strokes in cricket. Comparisons were between K-Nearest Neighbors (KNN), support vectors machines (SVMs) and CNNs as to which achieved the highest accuracy rating.

The aim of the feature extraction phase in Moodley and van der Haar's work focused on depicting a pattern as a vector of features. Feature extraction was achieved through two algorithms, Orientated FAST and rotated BRIEF (ORB) and HoG. The feature extraction techniques were applied to the KNN and SVM models.

The CNN model implemented, AlexNet architecture, requiring minimal pre-processing and constant image size at the input layer, dimensionality reduction made this feasible. The Convolutional network implemented a fully connected layer, max-pooling and a softmax layer to calculate the class of the cricket stroke.

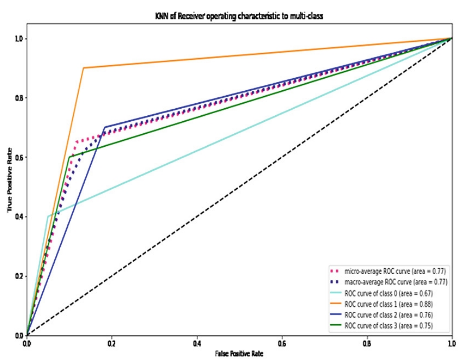

Figure 3: ROC KNN

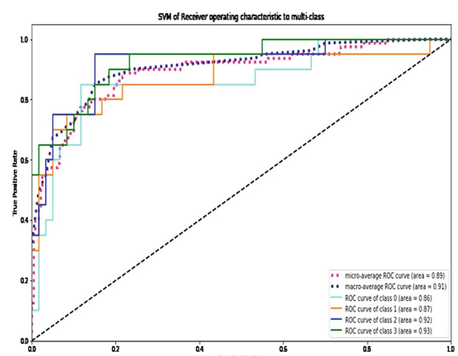

Figure 4: ROC SVM

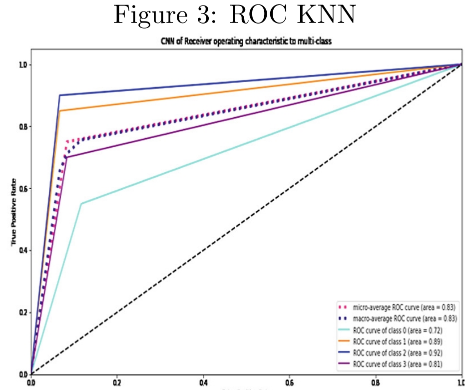

Figure 5: ROC CNN

The receiver operating characteristic (ROC) curves [5]of each model in Figure 3 to 5 show that the SVM model (Figure 4) achieves the highest micro and macro average of 89% and 91% respectively, using the three classes of cricket strokes. The false-positive rate (x-axis) vs true positive rate (y-axis) were the basis of the ROC curves. Overall the F1-score average of the CNN model was the deciding factor on the CNN model being the most accurate algorithm for stroke recognition Moodley and van der Haar (2020)

Contrasting the three models KNN, SVM, and CNN, the study found that the CNN model outperformed the other two models in all metric scores (Accuracy, precision, recall and F1-score), 74% for accuracy and precision and 75% for both recall and F1-score, respectively. The next closest model was the SVM obtaining an F1-score of 71%.

As the golfer is a vital aspect to extract and track from the video frames looking into Seemanthini and Manjunath (2018)'s study of human detection and tracking for action recognition is worthwhile analysing. Seemanthini and Manjunath emphasis is on detection of human object tracking while also incorporating difficult background conditions.

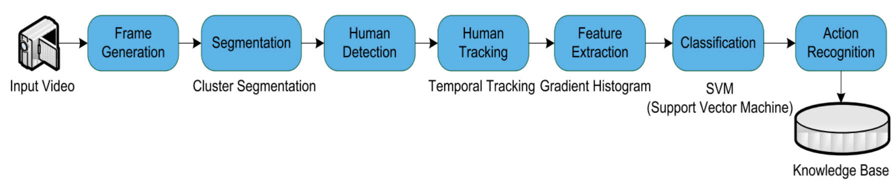

Figure 6: Human object recognition proposed architecture (Seemanthini and Manjunath, 2018)

Seemanthini and Manjunath's methodology as seen in Figure 6 first focuses on frame segmentation to extract humans from the background. To implement the segmentation, hierarchical graph-based segmentation containing two vital levels. A part level retrieval, ensuring colour-consistent clusters on the input

[5]ROC graphs have become increasing popular in the machine learning community as they enable users to visualise, organise and select classifiers based on their performance (Fawcett, 2006).

frames (Seemanthini and Manjunath, 2018). The colour-consistent clusters are then passed on to the object level for separation Seemanthini and Manjunath (2018). Using this technique, input frames are divided up into several colour consistent clusters.

The next phase is human tracking, following a humans path using a temporal based tracking model (Seemanthini and Manjunath, 2018). By implementing segmentation on each frame of the video dataset, human tracking can be achieved.

Feature extraction was the next hurdle in Seemanthini and Manjunath (2018)'s study. The model can compute gradients for both x and y directions by implementing the histogram gradient technique and then uses this information to extract patterns (Seemanthini and Manjunath, 2018).

The SVM then classifies the extracted patterns, and from this action recognition can take place for individual objects. Actions picked up in this study include gathering, walking, talking, etc... (Seemanthini and Manjunath, 2018).

The results of Seemanthini and Manjunath (2018)'s study found the model had an 89.59% accuracy at tracking humans.

Particular focus on the feature extraction implemented in Seemanthini and Manjunath's study could once again come in handy when tracking a golfers swing sequence and head movement.

3 Critical Analysis of Literature Review

Analysing the related studies, one can see that there are a lot of different aspects to cover. Image processing, as explained in Section 2.1, has proven successful, but this method is outdated. The most recent simple machine learning approaches tend to combine feature extraction techniques such as ORB and HoG with models such as KNN and SVM as explained in Section 2.2.3.

Feature extraction is a vital aspect as this is the foundation of the machine learning architecture. However, the implementation of lightweight CNNs with minimal image processing techniques has proved successful in Section 2.2.1. The minimal image processing techniques involved when implementing the CNN included resizing using bilinear interpolation and normalisation used on the ImageNet database, which is relatively simple to implement.

Moodley and van der Haar's stroke recognition model proved that simple machine learning algorithms such as KNN and SVMs combined with powerful feature extraction, i.e. HoG and ORB techniques, can achieve very close results to lightweight CNNs as discussed in Section 2.2.3.

Are state-of-the-art models such as YOLO or R-CNN capable of achieving better results or performing faster than simple CNNs that classifies on an entire image or SVM classifiers combined with image preprocessing techniques such as HoG or ORB? The study proposed by Shinde et al. on fast action recognition using YOLOv2 (Redmon and Farhadi, 2017) proved successful over other state-of-the-art models such as Mukherjee et al.'s SVM with leave-one-out cross-validation scheme, with both models performing classification on the LIRIS dataset (Wolf et al., 2014).

Overall, simple CNNs, YOLO or R-CNN seem to be the most likely candidates to achieving the best outcome for head and swing synchronisation during the golf swing. Therefore, if CNNs are implemented, the frame's foreground requires focusing purely on the golfer with minimal background variation. If there tends to be a lot of variation in the background and foreground, as is the case with the changing terrain on golf courses and the changing angle of viewing the golf swing, then use state-of-the-art models YOLO or R-CNN may be warranted. Fast R-CNN explained in Section 2.2.2 could be used to narrow down the RoI, which would be the golfer and golf club.

4 Plan of Action

4.1 Research Methodology Overview

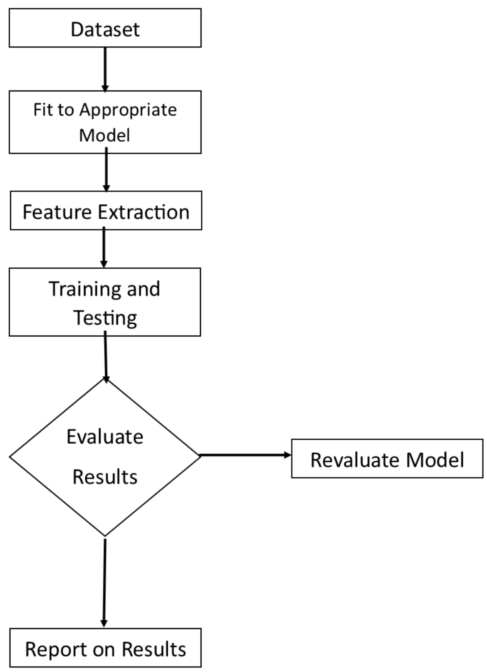

Figure 7: Overview of preliminary computer vision methodology.

The plan of action is summed up in Figure 7 expanding on each point below:

- **Dataset:** Acquiring a dataset of videos will be crucial to training and testing the model. In the initial phases of designing the model, a basic dataset will be constructed using a combination of datasets discussed in Thomas et al. (2017) containing golf swings. The GolfDB dataset (McNally et al., 2019) consisting of a combination of slow-motion and real-time golf swing videos will be incorporated into the model once the model's basic structure has been implemented successfully.

- **Fit to Appropriate Model:** Implementation of a state-of-the-art model seems to be the most viable option. The various models being considered are CNN's, YOLO or Fast R-CNN.

- **Feature Extraction:** Most of the feature extraction will be handled by the selected CNN approach. Image processing may be incorporated when cropping and resizing video frames is necessary to narrow down the region of interest, which would be the golfer and golf club.

- **Training and Testing:** A crucial step for accurate modelling, where all the action happens. Fine-tuning and applying the appropriate parameters to get the most out of the model and the given dataset to achieve the highest possible results from the model will be implemented during training. Possible results indicators will be based on the accuracy, recall and the F1-score of the model.

- **Evaluate Model:** Possible results indicators will be based on the accuracy, recall and the F1-score of the model. These indicators will be used to evaluate the model and areas of improvement at this stage, whether it is changing the model itself or adjusting the training and testing parameters.

- **Report on Results:** Findings of the model results will be discussed here. What aspects of the model performed well and possible shortcomings. The findings will also be compared to the results achieved in McNally et al. and Ko and Pan studies.

4.2 Timeline

Deadline	To Do
2 April 2021	Draft proposal.
6-20 April 2021	Seminar Series 1 presenting draft proposals.
8 April 2021	Completion of image processing task set to gain practical competency.
7 April 2021	Review feedback from seminar series 1 and apply relevant changes.
23 April 2021	Final proposal.
31 April 2021	Be able to extract the required information from video frames or images. i.e. golf club, arm and head.
1 May 2021	Start applying machine learning knowledge gained to working towards developing a machine learning model as machine learning course will be nearing completion.
20 May 2021	Datasets compiled
21 May 2021	Literature review submission.
31 May 2021	Suitable machine learning model selected by this stage.
15 June 2021	Final chance to change model.
15 July 2021	Achieve required feature extraction
31 July 2021	Working machine learning model, may still need some adapting at this stage. Testing will be done at this stage as well.
9-11 August 2021	Seminar Series 2: Presenting Project Progress.
30 August 2021	Completion of testing phase. Finalisation of technical project work.
20 September 2021	Progress report for supervisor.
4 October 2021	First full draft thesis.
11 October 2021	Submit short ACM-style paper.
18-20 October 2021	Seminar Series 3: Final Oral Presentations.
29 October 2021	Final Project Submission.
17-22 November 2021	Oral Examination.
3 December 2021	Corrected short ACM-style paper Corrected project

References

John Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. ISSN 01628828. doi: 10.1109/TPAMI.1986.4767851.

J. M. Cooper, B. T. Bates, J. Bedi, and J. Scheuchenzuber. Kinematic and kinetic analysis of the golf swing. *Biomechanics IV*, pages 298–305, 1974. doi: 10.1007/978-1-349-02612-8_44.

Hua Cui and Jie Bai. A new hyperparameters optimization method for convolutional neural networks. *Pattern Recognition Letters*, 125:828–834, 2019. ISSN 01678655. doi: 10.1016/j.patrec.2019.02.009.

Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. IEEE, 2009.

Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ISSN 01678655. doi: 10.1016/j.patrec.2005.10.010.

Nicolas Gehrig, Vincent Lepetit, and Pascal Fua. Visual golf club tracking for enhanced swing analysis. In *British Machine Vision Conference (BMVC)*, pages 1–10, 2003.

Ross Girshick. Fast R-CNN. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Pro-*

ceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), pages 580–587, 2014.

Paul Glazier and Peter Lamp. Golf Science: Optimum Performance From Tee to Green. In *Golf Science: Optimum Performance From Tee to Green*, chapter The swing. University of Chicago Press, 2013. ISBN 9780226001272.

Rafael Gonzalez. *Digital image processing.* Prentice Hall, Upper Saddle River, N.J, 2008. ISBN 978-0-13-168728-8.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 37(9):1904–1916, 2015.

Kyeong Ri Ko and Sung Bum Pan. CNN and bi-LSTM based 3D golf swing analysis by frontal swing sequence images. *Multimedia Tools and Applications*, pages 8957–8972, 2020. ISSN 15737721. doi: 10.1007/s11042-020-10096-0.

Yann LeCun, Koray Kavukcuoglu, and Clément Farabet. Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 253–256. IEEE, 2010.

Wei-Meng Lee. *Python Machine Learning.* John Wiley & Sons, Incorporated, Indianapolis, 2019. ISBN 9781119545637.

William McNally, Kanav Vats, Tyler Pinto, Chris Dulhanty, John McPhee, and Alexander Wong. GolfDB: A video database for golf swing sequencing. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2019-June:2553–2562, 2019. ISSN 21607516. doi: 10.1109/CVPRW.2019.00311.

Tevin Moodley and Dustin van der Haar. Cricket stroke recognition using computer vision methods. In *Information Science and Applications*, pages 171–181. Springer, 2020.

Snehasis Mukherjee, Apurbaa Mallik, and Dipti Prasad Mukherjee. Human action recognition using dominant motion pattern. In *International Conference on Computer Vision Systems*, pages 477–487. Springer, 2015.

Keiron O'Shea and Ryan Nash. An Introduction to Convolutional Neural Networks. *arXiv preprint arXiv:1511.08458*, pages 1–11, 2015. URL http://arxiv.org/abs/1511.08458.

Maria MP Petrou and Costas Petrou. *Image processing: the fundamentals*. John Wiley & Sons, 2010.

Joseph Redmon and Ali Farhadi. YOLO9000: Better, faster, stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017-January:7263–7271, 2017. doi: 10.1109/CVPR.2017.690.

Jürgen Schmidhuber. Deep Learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015. ISSN 18792782. doi: 10.1016/j.neunet.2014.09.003.

Mike Schuster and Kuldip Paliwal. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*, 45:2673 – 2681, 12 1997. doi: 10.1109/78.650093.

K. Seemanthini and S. S. Manjunath. Human detection and tracking using hog for action recognition. *Procedia Computer Science*, 132:1317–1326, 2018. ISSN 18770509. doi: 10.1016/j.procs.2018.05.048.

Shubham Shinde, Ashwin Kothari, and Vikram Gupta. Yolo based human action recognition and localization. *Procedia Computer Science*, 133:831–838, 2018. ISSN 18770509. doi: 10.1016/j.procs.2018.07.112.

Aimée Smith, Jonathan Roberts, Eric Wallace, and Stephanie Forrester. Professional golf coaches' perceptions of the key technical parameters in the golf swing. In *Procedia Engineering*, volume 34, pages 224–229. Elsevier Ltd, Jan 2012. doi: 10.1016/j.proeng.2012.04.039.

Graham Thomas, Rikke Gade, Thomas B. Moeslund, Peter Carr, and Adrian Hilton. Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159:3–18, Jun 2017. ISSN 1090235X. doi: 10.1016/j.cviu.2017.04.011.

Christian Wolf, Eric Lombardi, Julien Mille, Oya Celiktutan, Mingyuan Jiu, Emre Dogan, Gonen Eren, Moez Baccouche, Emmanuel Dellandréa, Charles Edmond Bichot, Christophe Garcia, and Bülent Sankur. Evaluation of video activity localizations integrating quality and quantity measurements. *Computer Vision and Image Understanding*, 127:14–30, 10 2014. ISSN 1090235X. doi: 10.1016/j.cviu.2014.06.014.