The Automation of Computer Vision Applications for Real-Time Combat Sports Video Analysis

Evan Quinn and Niall Corcoran
Technological University of the Shannon: Midlands Midwest, Moylish Campus, Co. Limerick, Ireland
evan.quinn@tus.ie

Abstract: This study examines the potential applications of Human Action Recognition (HAR) in combat sports and aims to develop a prototype automation client that examines a video of a combat sports competition or training session and accurately classifies human movements. Computer Vision (CV) architectures that examine real-time video data streams are being investigated by integrating Deep Learning architectures into client-server systems for data storage and analysis using customised algorithms. The development of the automation client for training and deploying CV robots to watch and track specific chains of human actions is a central component of the project. Categorising specific chains of human actions allows for the comparison of multiple athletes' techniques as well as the identification of potential areas for improvement based on posture, accuracy, and other technical details, which can be used as an aid to improve athlete efficiency. The automation client will also be developed for the purpose of scoring, with a focus on the automation of the CV model to analyse and score a competition using a specific ruleset. The model will be validated by comparing performance and accuracy to that of combat sports experts. The primary research domains are CV, automation, robotics, combat sports, and decision science. Decision science is a set of quantitative techniques used to assist people to make decisions. The creation of a new automation client may contribute to the development of more efficient machine learning and CV applications in areas such as process efficiency, which improves user experience, workload management to reduce wait times, and run-time optimisation. This study found that real-time object detection and tracking can be combined with real-time pose estimation to generate performance statistics from a combat sports athlete's movements in a video.

Keywords: computer vision, real-time, human action recognition, decision science, combat sports, object detection and tracking

1. Introduction

Artificial Intelligence (AI) is increasingly used in sports to analyse individual and team performance, improve spectator experiences, prevent injuries, establish betting odds, and help with officiating. This study focuses on using AI to improve the analysis and scoring processes in combat sports. Combat sports are competitive contact sports that usually involve one-to-one contact using a certain set of rules that determine the outcome, and include boxing, kickboxing, karate, Jiu-Jitsu, and Mixed Martial Arts (MMA). Martial arts are organised methods of teaching people how to fight and several martial arts are now considered combat sports. According to Davis, Wittekind and Beneke (2013), human error has hampered combat sports scoring. However, recent technological advances have enabled more precise data tracking that may help with this problem (Ishac and Eager, 2021), although each combat sport presents its own set of difficulties. For example, boxing has several scoring regulations that govern the sport to ensure fairness, safety, and competition for the fighters, and some organisations score competitions using a variety of scoring techniques.

The goal of this study is to investigate the automation capabilities of computer vision (CV) architectures for scoring combat sports and to generate human performance statistics based on observations, such as what a judge outside the competition area would do. Real-time object detection and Human Action Recognition (HAR) models are being investigated to quantitatively score video data segments of combat sports competitions and classify agents according to certain characteristics, such as apparel colour. The main objective is to create advanced CV systems for analysing and scoring combat sports, including boxing, kickboxing, and MMA. The scoring systems in certain combat sports are more complex. For example, there are two ways to gain an advantage over an opponent in MMA: creating kinetic energy by striking, (kicking, with knees, elbows, and fists), and generating isometric tension, which is used for strangulation and joint manipulation, and the scoring system is based on a combination of these factors. Combat sports experts must be used to validate CV system results to ensure they are accurate and not skewed. The purpose of HAR is to identify human movement in a video sequence and determine its duration. However, Host and Ivašić-Kos (2022) claim that using HAR to detect and recognise players during training matches, competition matches, and warmups in sports such as volleyball, basketball, soccer, and tennis, can be problematic. This study is structured as a HAR and object-tracking problem.

The study employs several CV techniques to classify and track baseline actions in real time such as punching, kicking, and inactivity. Combat sports are complex, necessitating the use of advanced video analysis systems to ensure accuracy (Krabben, Orth and van der Kamp, 2019). The primary objectives are to save time and reduce human error by automating data generation, to automate CV models that accurately score and analyse combat sports videos, and to create a large dataset of labelled combat sports videos that distinguish between various sports, strategies, techniques, and physical characteristics.

The following research questions have been developed for the study:

- RQ1: What technologies are required to generate human performance statistics from combat sports videos for improved coaching applications?
- RQ2: How can multiple athletes be tracked in real time using video data from various camera types?
- RQ3: How can the actions performed by an athlete be classified in real time using computer vision technologies?
- RQ4: How can human performance statistics be used to visualise a combat sports scenario?
- RQ5: What is required to manage real-time human performance statistics derived from videos?

Null and alternate hypotheses have been developed to examine the efficacy of CV automation systems in the context of combat sports:

> *H(0): The development of a real-time CV automation system restricts the generation of combat sports human performance statistics.*
>
> *H(1): The development of a real-time CV automation system enables the generation of combat sports human performance statistics.*

2. Literature review

Expert judges and statisticians are required in combat sports to objectively identify the athletes who are winning a competition, but this process can be automated thanks to advancements in wearable inertial sensors and CV. Much of the current research is centred on the potential applications of wearable sensors in combat sports. For example, Ishac and Eager (2021) use a combination of vision and inertial sensors to assess martial arts punching, focusing on strike velocity, impulse, momentum, and impact force. De Leonardis et al. (2018) perform HAR using wearable sensors and Saponara (2017) created a wearable sensor for measuring a combat sports athlete's performance. The CV field is expanding at a rapid rate due to the abundance of available visual data (Jaiswal et al., 2020) and CV is now being used to monitor athletic performance and automate statistical analysis (Host and Ivašić-Kos, 2022). According to Pang et al. (2022), CV can bring several benefits to combat sports, including more scientific training, more effective technical and tactical analysis, and the generation of performance metrics and statistics, and they noted that there are two primary methods for evaluating human action qualities, direct comparison and similarity calculations. Cardino, Chua and Llaga (2022) recognise that boxing scoring is subjective and that many factors can influence biases. They use YOLOv5-based models to classify human movements and generate scorecards, which are statistically compared to ground truth data verified by certified boxing judges. You Only Look Once (YOLO) is a simplified, one-step process for object detection and classification (Redmon et al., 2016).

Echeverria and Santos (2021) implement punch anticipation in karate using CV and conclude that using CV techniques to model and interpret karate body movements represents significant advances in HAR research for tracking combat sports. Van Zandycke et al. (2022) discuss the recent development of deep learning (DL) applied to CV and conclude that sports video understanding has gained a lot of attention, providing much richer information for both sports consumers and leagues. The datasets used in the study include high-resolution raw images, camera parameters, and high-quality annotations. When working with large datasets, it is critical to find high-resolution raw images that can be enhanced with best practices such as cropping, flipping, rotation, translation, brightness, contrast, colour augmentation, and saturation. Liu and Liu (2021) observe that an important goal of CV research is to give computers human-like cognitive abilities so that the computer can recognise the environment in the visual field, understand the content of emotions, and take appropriate actions.

2.1 Object tracking in video

Object tracking is a DL process in which an algorithm tracks the movement of an object. It is essentially the task of estimating or predicting the positions of moving objects in a video while also considering other, relevant information. Object tracking is usually preceded by object detection. Fernández et al. (2021) present TrafficSensor, a system that employs DL techniques for automatic vehicle tracking and classification on highways using a calibrated and fixed camera. TrafficSensor accurately detects and classifies the objects within images using various versions of YOLO. Kalake et al. (2022) propose a paradigm aimed at eliminating object tracking difficulties by enhancing the detection quality rate through the combination of a convolutional neural network (CNN) and a histogram of oriented gradient (HOG) descriptor. Jiang et al. (2022) propose a novel method to continuously track several mice and individual parts without requiring any specific tagging. Durve et al. (2022) examine the YOLOv5 object detector and DeepSORT object tracker for tracking droplets in microfluidic experiments. Yu et al. (2022) combine detection and embedding to train object detectors to track in the absence of fully annotated videos. Xu et al. (2022) develop a more efficient and effective solution to clean, stabilise, and label frames during training to avoid noisy, blurry, or unlabelled background frames. Muller et al. (2018) present TrackingNet, the first large-scale dataset and benchmark for object tracking in the wild. In addition, a benchmark of 500 new videos with distributions resembling the training dataset was created by TrackingNet. TrackingNet provides a fair benchmark for the future development of object trackers by encrypting the test set annotation, providing an online evaluation server, and evaluating more than 20 trackers. Ahmadyan et al. (2021) introduce the Objectron dataset to advance the state of the art in 3D object detection and promote new research, such as 3D object tracking, view synthesis, and improved 3D shape representation. Tokmakov et al. (2021) present an end-to-end trainable approach for joint object detection and tracking that is capable of object permanence and approximate object localization in the presence of full occlusions.

2.2 Pose estimation in video

Pose estimation is a CV technique that predicts and tracks a person's or object's location. This is accomplished by examining a given person's or object's pose and orientation. Bridgeman et al. (2019) present an approach to multi-person 3D pose estimation and tracking from multi-view video in sports and conclude that generating 3D pose estimations takes a long time and limits its applicability in sports. However, their proposed method significantly outperforms other state-of-the-art methods in terms of speed. Zecha, Einfalt and Lienhart (2019), in investigating the automation of performance evaluation and motion dynamics prediction in sports, focus on aquatic training scenarios, where even novel pose estimators produce several types of orthogonal errors, including joint swaps and prediction outliers. Sengupta, Budvytis and Cipolla (2020) implement synthetic training data for accurate 3D pose estimation from RGB images. Li et al. (2021) use pose estimation for coaching baseball swings and propose a multi-residual module CNN-based method for athlete pose estimation in sports game videos. The results of several groups of comparative experiments show that the algorithm can better estimate human posture and has good performance in solving multi-person pose estimation in sports game videos (Guo, 2022). Zhang et al. (2017) introduce the Martial Arts, Dancing and Sports (MADS) dataset, which consists of challenging martial arts actions (Tai-chi and Karate), and sports actions (basketball, volleyball, football, rugby, tennis, and badminton). Wang (2022) proposes a CNN for athlete pose estimation in sports game videos. However, the method used only estimates the pose of a single athlete and cannot solve the pose estimation problem of multiple athletes. The intermediate supervision method is used to avoid the gradient vanishing problem of the CNN.

3. Research methodology

Combat sports observations are abstract, influenced by variables that are hard to quantify. The main reason for using quantitative research is to analyse a combat sports video sequence objectively and accurately in real time using CV technologies. Therefore, this study takes a pragmatic approach as it incorporates operational choices based on what will be most effective in providing answers to the research questions. High-quality data is essential to evaluate an algorithm’s performance, allowing for the development of more accurate algorithms. Secondary data sources are evaluated to ensure that the results are not misrepresented. Coding enables the following of important actions in the data, which is counted quantitatively. The goal is to create a narrative analysis of the video sequences using content analysis.

The primary goal is to confirm whether the findings obtained from both sources (researcher and external observer) are parallel, which is referred to as triangulation. Most of the video data were obtained from data sources ranging from online open-source datasets to videos collected from YouTube and self-generated videos to test whether deep learning models can be used in real-world practical situations. The performance of combat athletes will be evaluated and compared to the classifications generated by the CV models. The CV models will be examined for correlations in performance and accuracy across multiple combat sports domains, including Olympic Boxing, MMA, and Person vs Boxing-Bag. The data collected in this study is a mix of experiments and controlled observations. Descriptive statistical analysis will identify target groups in the video data to determine how many frames are mislabelled during experimentation. Relationship data reveals patterns between two or more variables.

3.1 Training and test sets

The image dataset was split into samples of combat sports classifications of Olympic Boxing, MMA, and Person vs Boxing-Bag (see Figure 1). The training sample for Olympic Boxing included over 800 samples of images that were used to train a YOLOv5 model. The training sample for MMA included over 1000 samples of the fighter class. The training sample for Person vs Boxing-Bag included over 1200 samples for the person class and 200 samples for the boxing-bag class. Data augmentation techniques were used on all datasets such as random rotation, flipping, and shifting. The test sample for Olympic Boxing was performed on three random Olympic Boxing videos that were not used to train the model. The test sample for MMA was performed on videos that were not used to train the model. The Person vs. Boxing-Bag test sample was performed using YouTube videos and data generated with a video camera. The Olympic Boxing test sets aim to demonstrate that object detection and tracking for other combat sports is possible, which is supported by MMA and Person vs Boxing-Bag samples.

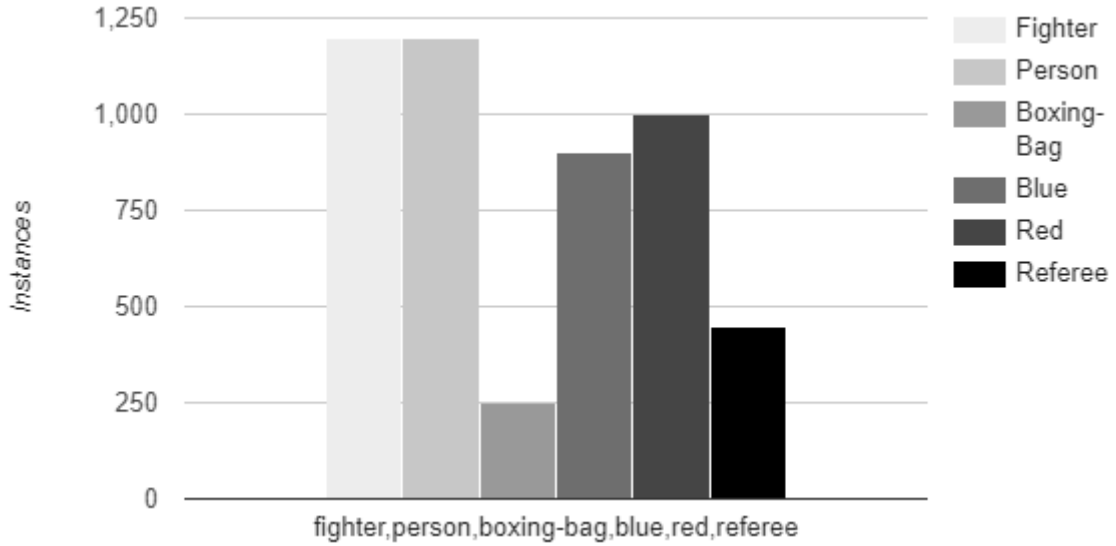

Figure 1: Dataset classifications used to train the YOLOv5 object detection and tracking models

3.2 Classifications with deep learning

YOLOv5 was used for the primary object detection and tracking framework for experiments. YOLOv5 is a family of object detection architectures and models pretrained on the MS COCO dataset (Ultralytics, 2022). The MS COCO dataset is a large-scale object detection, segmentation, and captioning dataset published by Microsoft. Machine Learning and CV engineers commonly use the COCO dataset for various computer vision projects. The YOLOv5m model was used in most experiments. When tracking combat sports classifications, the batch size and epochs were fine-tuned for performance and accuracy. The customized YOLOv5 model allowed for accurate object tracking and the ability to read and write large amounts of data in real time for processing. A Windows workstation equipped with an RTX 2080 was used for training and evaluation.

3.3 Evaluation

This quantitative study collects data to examine the research questions and hypotheses and analyses the data using descriptive and inferential statistics. The descriptive statistical methods include mean, median, standard deviation, and skewness of performance, accuracy, and training statistics for various architectures running in real time. The inferential statistics include correlation and regression analysis to examine the hypotheses. Primary data is generated through experiments. A precision-recall curve shows the relationship between precision (positive predictive value) and recall (sensitivity) for every possible cut-off. Precision (P) and Recall (R) are two of the most important metrics to look at when evaluating an imbalanced classification model (Davis and Goadrich, 2006). The precision-recall curve shows the trade-off between precision and recalls for different thresholds. Average precision ranges from the frequency of positive examples, 0.5 for balanced data to 1.0 for a perfect model. The F1 Curve is the weighted harmonic mean of P and R of a classifier, taking α=1 (F1 score). The P and R scores have the same importance. In most cases, higher confidence values and F1 scores are desirable. The F1 score is applicable for any point on the receiver operating characteristic (ROC) curve.

4. Results

4.1 Olympic Boxing

Most experiments were done on the Olympic Boxing dataset. The primary physical characteristics being classified are the colours of each athlete's clothing, so the three main categories tested are red, blue, and referee. For example, the blue fighter is easily distinguished from the red fighter, and the referee is dressed in white and black. Over 800 images from the blue and red classes were used to train the model. The disparity in the referee class was caused by the data sample consisting of more images with only the red and blue classes because the referee was out of the Field-of-View (FOV). The three classes are equally important, and they correctly classify 80% to 90% of the expected results. The mean average precision (mAP) is used to measure the performance of object detection models. Four primary video samples were used for the Olympic Boxing experiments. Sample One is the primary video used to train the YOLOv5 model, and it was also augmented to perform on other similar data sources. The test case is made up of 30 frames chosen at random from the video. The expected levels of confidence are in the upper 90% of the range, and each frame has been validated and is tracking the data as expected. The confidence percentage metrics for the 30 random frames in the four samples are shown in Table 1. The mean for the referee class is significantly lower than the mean for both the red and blue classes. The mean would be higher if the referee was present in more of the frames, but the referee was absent 12 times out of the 30 random frames, accounting for 40% of the total frames in Sample One. The referee was not visible in Sample Two or Sample Four. Sample Two met expectations and the only frame that could not be validated as a label was due to a low confidence value (0%). This problem could have been avoided by labelling more information about that frame. The confidence values are visible on the bounding boxes for the athletes as shown in Figure 2. In Sample Two, the blue class is not visible in one frame. If the confidence threshold was set to lower than 0.4, the blue class may have been recognised. To overcome this problem, a more extensive dataset is required. For Sample Three, two frames were labelled incorrectly by the model. The blue class is not visible in the camera in two of the frames and the red class is duplicated in one of the frames.

The results of the Olympic Boxing experiments are similar to the stated results by Cardino, Chua and Llaga (2022). All classes achieved 0.955 mAP @ 0.5. This means that all classes are predicted with 95.5% accuracy at a confidence threshold of 50%. The model performs accurately on a wide range of Olympic Boxing video samples and the PR curve is shown in Figure 3. Although the model was not trained on Samples Two, Three, or Four, the data produced was correctly classified in a high percentage of experiments. However, more detailed datasets are needed to account for a wide range of potential combat sports data variations. The F1 score for all classes in Sample One is 0.95 at a confidence threshold of 0.489. A higher confidence value and F1 score are usually preferred. The F1 curve for all samples is shown in Figure 3. The F1 score for experiments could be increased by focusing on generating a rich dataset. Techniques such as data augmentation, limited data training, and transfer learning can be investigated further to narrow down the highest performance techniques for tracking combat sports in real-time using CV architectures. For this research, the CV model was trained for performance on the video called Sample One and was not expected to perform as accurately as Samples Two, Three, and Four. In general, the experiments performed more accurately than was expected.

Table 1: Confidence metrics for Olympic Boxing samples

		Video Frames	Count (n)	Standard Deviation (s)	Sum (Σx)	Mean (x̄)	Variance (s2)
Sample One	Red	22044	30	1.393333883	2841	94.7	1.94137931
	Blue		30	3.673608893	2723	90.766667	13.4954023
	Referee		30	45.51043836	1644	54.8	2071.2
Sample Two	Red	21897	30	23.35532625	2270	75.666667	545.4712644
	Blue		30	5.854166411	2662	88.733333	34.27126437
	Referee		-	-	-	-	-
Sample Three	Red	7654	30	9.030217472	2694	89.8	81.54482759
	Blue		30	23.94149958	1760	58.666667	573.1954023
	Referee		30	4.470079328	2626	87.533333	19.9816092
Sample Four	Red	27075	30	0.647719252	2855	95.166667	0.41954023
	Blue		30	1.295882072	2823	94.1	1.679310345
	Referee		-	-	-	-	-

Figure 2: Sample One training data set

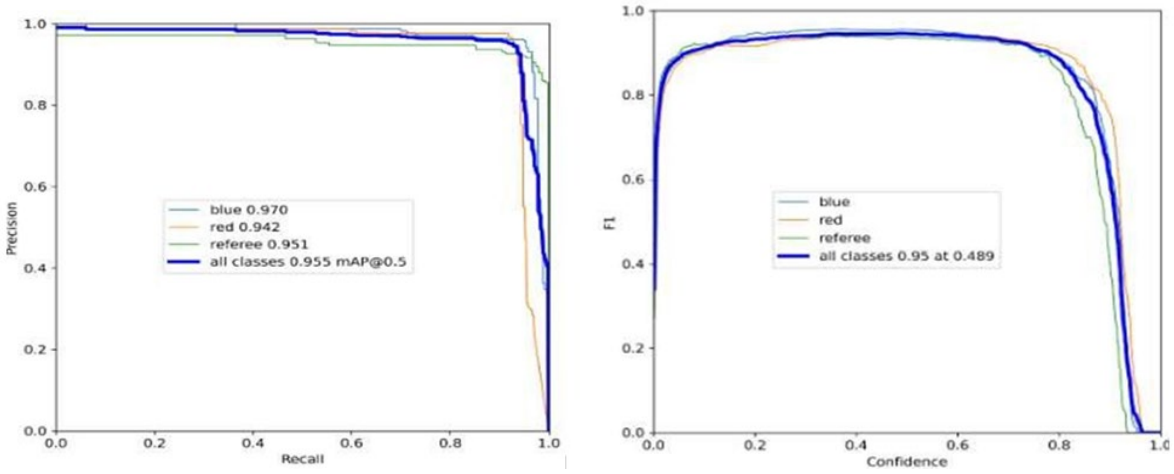

Figure 3: PR and F1 curves for Olympic Boxing samples

4.2 Person vs Boxing-Bag

In this experiment, the main categories tested for object detection and tracking and HAR are 'Person' and 'Boxing-Bag'. The potential for tracking a person with a video camera and counting the number of hits landed on a boxing bag was investigated. Figure 4 shows object tracking and hit detection. When the bounding box overlaps, the hit counter is flagged at true, or 1. If no overlap is present in the frame, the hit counter is flagged at false, or 0.

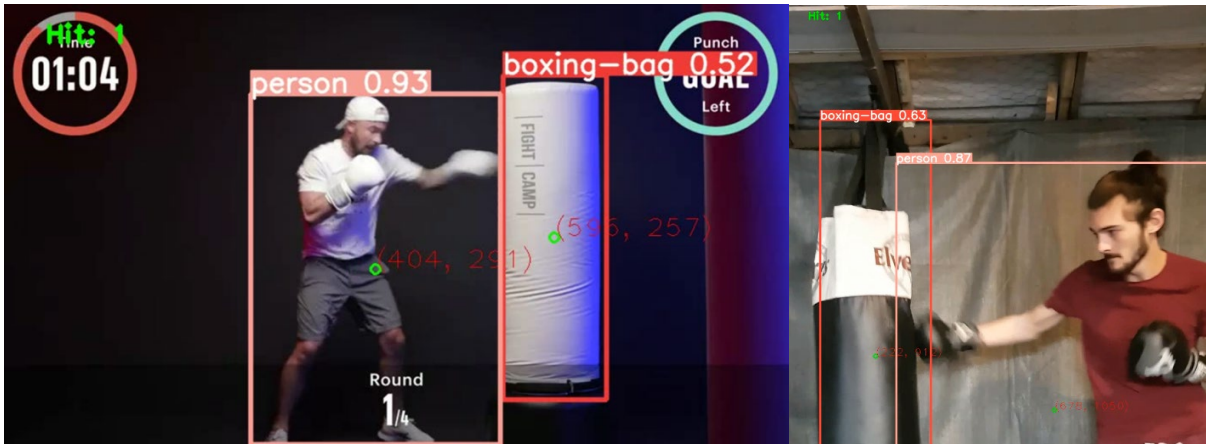

Figure 4: Person vs Boxing-Bag with object tracking and hit detection

To provide more advanced statistics, the pose estimation framework could be combined with real-time YOLOv5 and Intersection over Union (IoU) counters. The model aims to accurately track when the person and the boxing-bag classes overlap which is recognised as a hit. By combining the hit detection system and HAR, it is possible to identify and generate combat sports techniques like punch, kick, and in-activity (see Figure 5). The overlap is conditional; if one object overlaps another, actions can be assigned to the frame. The F1 score performs above expectations and accurately tracks both classes in real-time, and all classes have an F1 score of 0.99 at a confidence of 0.684. Hit or no hit can be combined with other HAR classifications, such as punch and kick, to generate performance statistics for combat sports techniques. More advanced combat sports statistics will be possible if the HAR model is trained to recognize increasingly advanced actions, such as breaking down punches to more detail such as jab, straight, and upper cut.

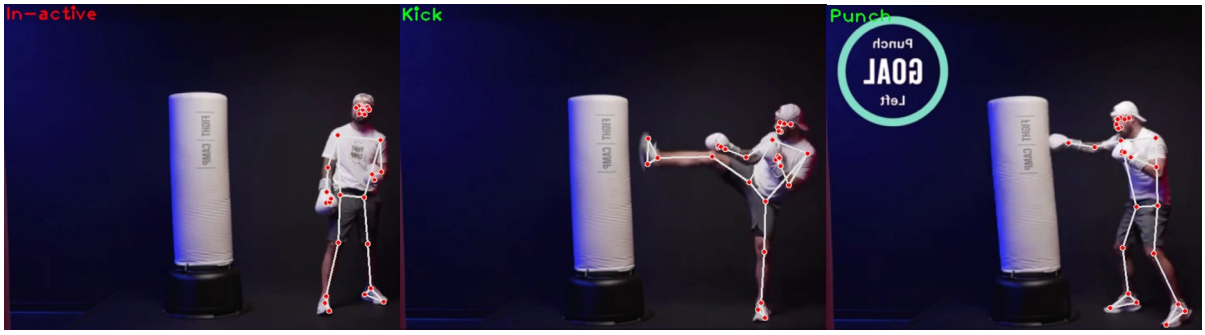

Figure 5: HAR samples with YouTube video to demonstrate various options, such as punch, kick, and inactivity

4.3 MMA

The goal of this experiment is to determine if it is possible to track and detect a fighter in an MMA competition. It was discovered through data sampling that the model's accuracy could be adjusted to ignore background information after training in a single class. The fighter class is trained with over 1000 different fighters. The MMA dataset's biggest limitation is its size and depth of information. The possibility of using HAR for MMA analysis was investigated. MMA requires advanced HAR systems to accurately describe human actions in a video file. Most experiments concentrate on 2D pose estimation. If a punch or kick in MMA can be tracked, and object detection and tracking are possible, scoring MMA competitions using CV applications appears likely when performance and accuracy are considered. Figure 6 illustrates how pose estimation key points are measured using angles for HAR.

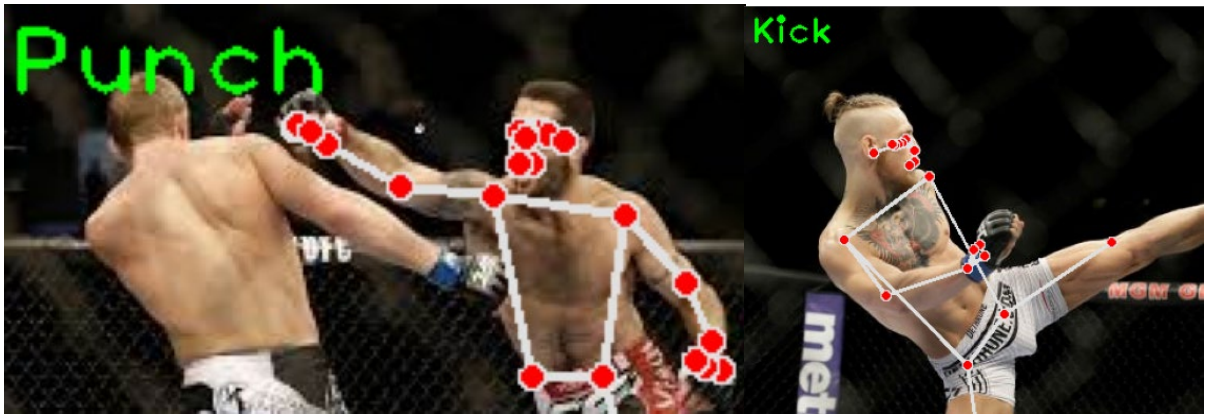

Figure 6: Samples of HAR for MMA

The dataset tracks fighters with high precision. The YOLOv5m model used for the experiment ran in real-time, and the object detection and tracking for the fighter class were accurate and outperformed expectations. The MMA training dataset is labelled simply with one class (fighter) and could be improved to perform a broader range of classifications such as referee, shorts, gloves, and so on. Figure 7 illustrates a HAR sample combined with the corresponding 3D pose estimation. To date, only preliminary experiments for 3D pose estimation have been carried out. Most of the research focuses on the possibility of 2D pose estimation as well as real-time object detection and tracking. 3D pose estimation could help with the analysis of a wide range of human movements.

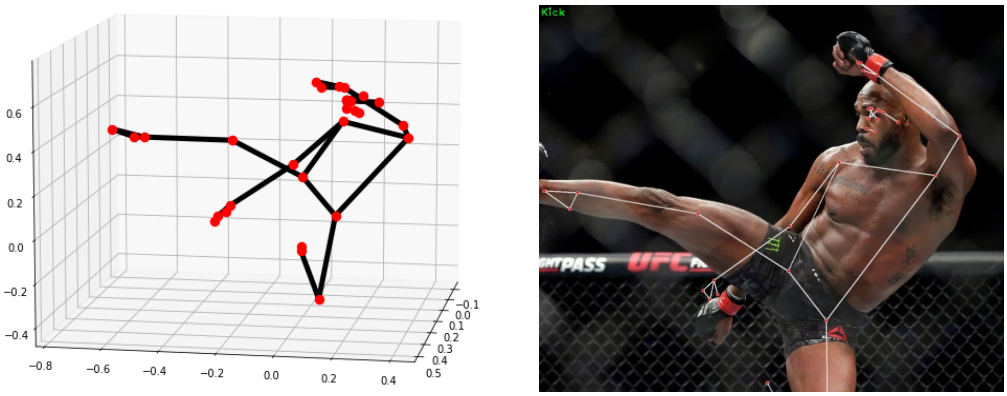

Figure 7: A HAR sample for MMA with 3D pose estimation

5. Discussion

The goal of conducting experiments on three separate combat sports samples was to demonstrate that different CV techniques work in a variety of settings. If an Olympic Boxing competition can be successfully tracked using CV, it suggests that MMA or Person vs Boxing-Bag can also be tracked. Large-scale video analysis of combat sports is possible by combining HAR, hit detection, and object detection and tracking, if an effort is put into creating a diverse dataset for HAR and object tracking. The experiments on Olympic Boxing were designed to demonstrate the feasibility of real-time object detection and tracking. The HAR experiments aim to show that detecting a punch, kick, or inactivity is possible for combat sports. Further research is required to improve the accuracy and reliability of the HAR classifications which will need to be compared to expert analysis in the field of combat sports to prove that the results are not biased. However, the depth of data required to generate more universal models that can be used to develop more advanced experiments is a problem that must be addressed. It is possible to track objects such as people, boxing bags, gloves, clothing, and so on. As a result, real-time combat sports video analysis is possible if CV models are trained for performance and accuracy. It was critical to manage the various resources as efficiently as possible due to the large amount of video data required to train high-performance CV models. The Olympic Boxing sample results show that CV architectures can be accurately detected and tracked in real-time for a referee and a red and blue fighter (see Figure 8). The study also examined MMA videos and followed fighters in an octagon, but more data is required to improve the results. The need to sort through large amounts of video samples to select videos that can be used to train and validate the performance of a CV model is a limitation.

Best practices and principles in data management are crucial for real-world data applications (Reno et al., 2018). For instance, large amounts of data are produced when deep learning models automatically identify patterns in data. Therefore, to manage the video data and quantitative data generated during analysis, each experiment needs to be examined, stored, and decomposed into individual use cases using coding systems related to each research question.

Figure 8: Examples of YOLOv5 and 2D multi-person pose estimation

6. Conclusion

This study discovered that real-time object detection and tracking can be combined with real-time pose estimation to generate statistics based on a combat sports athlete's movements in a video file. A modified version of YOLOv5 was used to track and classify various combat athletes in real time. All the research questions have been addressed through experiments and the hypothesis (H0) has been rejected. It appears feasible to use this CV model to recognize specific combat sports movements and generate statistics in real time. Therefore, it is possible to develop a real-time CV automation system to generate combat sports human performance statistics. Further research can use alternative object detection and tracking architectures, such as YOLOv7, to improve the accuracy and performance of the results.

References

Ahmadyan, A., Zhang, L., Ablavatski, A., Wei, J. and Grundmann, M. (2021) "Objectron: A Large Scale Dataset of Object-Centric Videos in the Wild With Pose Annotations", Paper read at IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 25 June.

Bridgeman, L., Volino, M., Guillemaut, J.-Y. and Hilton, A. (2019) "Multi-Person 3D Pose Estimation and Tracking in Sports", Paper read at IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Nashville, TN, USA, 19-25 June.

Cardino, P. C. C., Chua, J. L. T. and Llaga, J. R. R. R. (2022). Advance Scorecard for Boxing: Combat Sport Analysis with Deep Learning. Bachelor of Science Thesis, Gokongwei College of Engineering.

Davis, J. and Goadrich, M. (2006) "The Relationship Between Precision-Recall and ROC Curves", Association for Computing Machinery, Pittsburgh, Pennsylvania, USA.

Davis, P., Wittekind, A. and Beneke, R. (2013) "Amateur Boxing: Activity Profile of Winners and Losers", International Journal of Sports Physiology and Performance, Vol. 8, No. 1, pp 84–92.

De Leonardis, G., Rosati, S., Balestra, G., Agostini, V., Panero, E., Gastaldi, L. and Knaflitz, M. (2018) "Human Activity Recognition by Wearable Sensors: Comparison of Different Classifiers for Real-time Applications", Paper read at IEEE International Symposium on Medical Measurements and Applications, Rome, Italy, 11-13 June.

Durve, M., Tiribocchi, A., Bonaccorso, F., Montessori, A., Lauricella, M., Bogdan, M., Guzowski, J. and Succi, S. (2022) "DropTrack—Automatic Droplet Tracking with YOLOv5 and DeepSORT for Microfluidic Applications", Physics of Fluids, Vol. 34, No. 8, pp 1-24.

Echeverria, J. and Santos, O. C. (2021) "Punch Anticipation in a Karate Combat with Computer Vision", Paper read at UMAP '21: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, Utrecht, Netherlands, 21-25 June.

Fernández, J., Cañas, J. M., Fernández, V. and Paniego, S. (2021) "Robust Real-Time Traffic Surveillance with Deep Learning", Computational Intelligence and Neuroscience, Vol. 2021, No. pp 1-18.

Guo, X. (2022) "Research on Multiplayer Posture Estimation Technology of Sports Competition Video Based on Graph Neural Network Algorithm", Computational Intelligence and Neuroscience, Vol. 2022, No. 1, pp 1-12.

Host, K. and Ivašić-Kos, M. (2022) "An Overview of Human Action Recognition in Sports Based on Computer Vision", Heliyon, Vol. 8, No. 6, pp 1-25.

Ishac, K. and Eager, D. (2021) "Evaluating Martial Arts Punching Kinematics Using a Vision and Inertial Sensing System", Sensors, Vol. 21, No. 6, pp 1-25.

Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D. and Makedon, F. (2020) "A Survey on Contrastive Self-Supervised Learning", Technologies, Vol. 9, No. 1, pp 1-22.

Jiang, Z., Liu, Z., Chen, L., Tong, L., Zhang, X., Lan, X., Crookes, D., Yang, M.-H. and Zhou, H. (2022) "Detecting and Tracking of Multiple Mice Using Part Proposal Networks", IEEE Transactions on Neural Networks and Learning Systems, Vol. 1, No. 1, pp 1-15.

Kalake, L., Dong, Y., Wan, W. and Hou, L. (2022) "Enhancing Detection Quality Rate with a Combined HOG and CNN for Real-Time Multiple Object Tracking across Non-Overlapping Multiple Cameras", Sensors, Vol. 22, No. 6, pp 21-23.

Krabben, K., Orth, D. and van der Kamp, J. (2019) "Combat as an Interpersonal Synergy: an Ecological Dynamics Approach to Combat Sports", Sports Medicine, Vol. 49, No. 12, pp 1-12.

Li, Y. C., Chang, C. T., Cheng, C. C. and Huang, Y. L. (2021) "Baseball Swing Pose Estimation Using OpenPose", Paper read at IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI), Tianjin, China, 21-23 April.

Liu, S. and Liu, Y. (2021) "Application of Human Movement and Movement Scoring Technology in Computer Vision Feature in Sports Training", IETE Journal of Research, Vol. 1, No. 1, pp 1-7.

Muller, M., Bibi, A., Giancola, S., Alsubaihi, S. and Ghanem, B. (2018) "Trackingnet: A Large-Scale Dataset and Benchmark for Object Tracking in the Wild", University in King Abdullah University Of Science And Technology, Saudi Arabia, KAUST, University.

Pang, Y., Wang, Q., Zhang, C., Wang, M. and Wang, Y. (2022) "Analysis of Computer Vision Applied in Martial Arts", Paper read at 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 14-16 January.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) "You Only Look Once: Unified, Real-Time Object Detection", Paper read at IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 1 January.

Reno, V., Mosca, N., Marani, R., Nitti, M., D'Orazio, T. and Stella, E. (2018) "Convolutional Neural Networks Based Ball Detection in Tennis Games", Paper read at IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18-22 June.

Saponara, S. (2017) "Wearable Biometric Performance Measurement System for Combat Sports", IEEE Transactions on Instrumentation and Measurement, Vol. 66, No. 10, pp 1-11.

Sengupta, A., Budvytis, I. and Cipolla, R. (2020) "Synthetic Training for Accurate 3D Human Pose and Shape Estimation in the Wild", Department of Engineering, University of Cambridge, Cambridge, UK.

Tokmakov, P., Li, J., Burgard, W. and Gaidon, A. (2021) "Learning to Track with Object Permanence", Paper read at IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11 October.

Ultralytics. (2022). YOLOv5 [Online]: GitHub, https://github.com/ultralytics/yolov5.

Van Zandycke, G., Somers, V., Istasse, M., Del Don, C. and Zambrano, D. (2022) "DeepSportradar-v1: Computer Vision Dataset for Sports Understanding with High Quality Annotations", MMSports '22: 5th International ACM Workshop on Multimedia Content Analysis in Sports Vol. 5, No. 5, pp 1-8.

Wang, Q. (2022) "Application of Human Posture Recognition Based on the Convolutional Neural Network in Physical Training Guidance", Computational Intelligence and Neuroscience, Vol. 2022, No. 1, pp 1-11.

Xu, M., Fu, C.-Y., Li, Y., Ghanem, B., Perez-Rua, J.-M. and Xiang, T. (2022) "Negative Frames Matter in Egocentric Visual Query 2D Localization", KAUST University, Saudi Arabia.

Yu, S., Wu, G., Gu, C. and Fathy, M. E. (2022) "TDT: Teaching Detectors to Track without Fully Annotated Videos", Paper read at CVPR 2022 Workshop: Workshop on Learning with Limited Labelled Data for Image and Video Understanding (L3D-IVU), New Orleans, Louisiana, USA, 21 - 24 June

Zecha, D., Einfalt, M. and Lienhart, R. (2019) "Refining Joint Locations for Human Pose Tracking in Sports Videos", Paper read at IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16-17 June.

Zhang, W., Liu, Z., Zhou, L., Leung, H. and Chan, A. B. (2017) "Martial Arts, Dancing and Sports Dataset: A Challenging Stereo and Multi-View Dataset for 3D Human Pose Estimation", Image and Vision Computing, Vol. 61, No. 2017, pp 22-39.