ChatPPG: Computational Analysis and Statistics of Table Tennis Games

GuangLiang Yang

A project report submitted to the Auckland University of Technology

in partial fulfillment of the requirements for the degree of

Master of Computer and Information Sciences (MCIS)

2024

School of Engineering, Computer & Mathematical Sciences

**Abstract**

This study presents an innovative integrated framework that combines multi-camera 3D reconstruction, deep learning models (YOLO and Transformer), and Low-Rank Adaptation (LoRA) for large language models (LLMs) to enable real-time analysis and evaluation of table tennis games. By leveraging precise ball detection, trajectory tracking, and serve foul detection, the system automates the entire process from data acquisition to interactive feedback. The multi-camera setup facilitates accurate 3D trajectory reconstruction to analyze ball speed, trajectory turning points, landing distributions, and player movements. Deep learning models significantly enhance small object detection accuracy (achieving an mAP@0.5 of 86.87%), while the Transformer model identifies key trajectory points such as the throw, highest, and hit points with an accuracy of 93%. Furthermore, by integrating LLMs with computer vision, the ChatPPG system enables semantic interpretation of data and interactive guidance functionalities. Fine-tuned using LoRA on a domain-specific dataset, the LLM achieved a Q/A accuracy of 92.3%. User feedback highlights the system’s practical value, with tactical suggestions and training plans scoring 9.3/10 and 8.9/10, respectively. This research addresses the gap in traditional sports data analysis tools, which often lack semantic understanding and interactive capabilities. By bridging data analysis with decision support, it sets a new standard for AI applications in high-speed sports analytics..

**Keywords**: LLMs, CV, Table Tennis, LoRA, Prompt Engineering, Function Calling, Model Quantization, Transformer, deep learning, Multi-Camera System, 3D Trajectory Analysis, Small Object Detection,

# Table of Contents

[Chapter 1 Introduction 1](#_Toc146331042)

[1.1 Background and Motivation 2](#_Toc146331043)

[1.2 Research Questions 3](#_Toc146331044)

[1.3 Contributions 4](#_Toc146331045)

[1.4 Objectives of This Report 5](#_Toc146331046)

[1.5 Structure of This Report 6](#_Toc146331047)

[Chapter 2 Literature Review 7](#_Toc146331048)

[2.1 Introduction 8](#_Toc146331049)

[2.2 Human Pose Estimation (HPE) 8](#_Toc146331050)

[2.3 Deep Learning Model 12](#_Toc146331051)

[2.4 HPE with Deep Learning 19](#_Toc146331052)

[Chapter 3 Methodology 22](#_Toc146331053)

[3.1 Transformer 23](#_Toc146331054)

[3.2 Ablation Experiment 27](#_Toc146331055)

[3.3 Key-Point Skeleton Model 29](#_Toc146331056)

[3.4 Posture Evaluation 31](#_Toc146331057)

[3.5 Billiards Player Striking Posture Dataset 33](#_Toc146331058)

[Chapter 4 Results 36](#_Toc146331059)

[4.1 Real-time Pose Comparison 37](#_Toc146331060)

[4.2 Posture Integrity Evaluation 40](#_Toc146331061)

[4.3 Training Transformer Model 41](#_Toc146331062)

[4.4 Ablation Experiment 44](#_Toc146331063)

[4.5 Comparative Analysis of Transformer with RNN and LSTM 49](#_Toc146331064)

[4.6 Sliding Window 52](#_Toc146331065)

[4.7 Real-time Billiard Player Pose Analysis 53](#_Toc146331066)

[Chapter 5 Analysis and Discussions 56](#_Toc146331067)

[5.1 Analysis 57](#_Toc146331068)

[5.2 Discussions 57](#_Toc146331069)

[Chapter 6 Conclusion and Future Work 59](#_Toc146331070)

[6.1 Conclusion 60](#_Toc146331071)

[6.2 Future Work 60](#_Toc146331072)

[References 61](#_Toc146331073)

**List of Figures**

Figure 2.1 BlazePose extracted key point skeleton ….…………………………………11

Figure 2.2 External schematic of the LSTM module ...…………………………………13

Figure 2.3 Internal structure of the LSTM………………………………………………15

Figure 2.4 The structure of Transformer model…………………………………………17

Figure 3.1 Transformer model structure..……………………………………………….24

Figure 4.1 Extracted standardized bone models and joint angles……………………......37

Figure 4.2 Skeletal modeling and angle analysis for real-time detection………...……...38

Figure 4.3 Comparison of real-time skeleton with standard skeleton.………………….39

Figure 4.4 Scoring system based on joint angle calculation…….……………….…...…40

Figure 4.5 Transformer model accuracy, loss, and F1 score plots..…………………...…43

Figure 4.6 Transformer training and validation response speed.…………………….....44

Figure 4.7 512 Model dimension accuracy and speed.……………………….………....47

Figure 4.8 The loss curves for 6 layers and 10 layers……………………………………48

Figure 4.9 Accuracy and loss graphs for the RNN.…………………………...…………51

Figure 4.10 Sliding window stabilization pose model.……………………………….…53

Figure 4.11 A system for analyzing the striking posture of billiard players.……….……54

**List of Tables**

Table 3.1 List of ablation experiments…………………………………………….……28

Table 4.1 Results of activation function ablation experiments.…………………………45

Table 4.2 Results of model structure ablation experiments………………...…..….…….45

Table 4.3 The results of the multiple attention heads ablation experiment.…………..….46

Table 4.4 Model dimension ablation experiment results……………………….……….46

Table 4.5 Results of ablation experiments…………..…………………….……….…...48

Table 4.6 Optimal model structure and parameters..……………………………….…...49

Table 4.7 Transformer, RNN and LSTM optimal model parameters…………..…….….49

**Attestation of Authorship**

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature: Date: 20 December 2024

A logo with black text

Description automatically generated

**Acknowledgment**

Fisrt of all, I would like to thank my parents for their finanical support. Owing to the unselfish and generous sponsor from them, I have this invaluable opportunity to complete my Master’s study with the Auckland University of Technology (AUT), New Zealand.

I would also like to express my deepest gratitude to my supervisor Wei Qi Yan. In this study, he not only provided me with professional knowledge support and careful guidance, but also helped me enrich my learning experience. I believe I could not complete my study without Dr Yan’s supervision and instructions. Meanwhile, I also appreciate the secondary supervisor Minh Nguyen and the school administrators of AUT for their invaluable guidance.

GuangLiang Yang

Auckland, New Zealand

December 2024

# Chapter 1 Introduction

*This chapter is composed of five parts: The first part introduces the background and motivations, the second part includes the research question, followed by the contributions, objectives, and structure of this report.*

## Background and Motivation

Table tennis is a fast-paced, precision-driven sport that demands athletes to execute complex technical actions under intense pressure. With the sport’s evolution, there is a growing demand for accurate and automated systems capable of analyzing ball trajectories, player movements, and serve compliance (Doe & Johnson, 2020; Zhang et al., 2022). Traditional methods, such as manual referee judgments or single-camera video analysis, often fall short in delivering the precision and consistency required for high-stakes competitions. These shortcomings not only impact match fairness but also limit the actionable feedback available to players and coaches for performance improvement.

Recent breakthroughs in computer vision (CV) and deep learning have revolutionized sports analytics. YOLO (You Only Look Once) models have proven highly effective in detecting small, fast-moving objects like table tennis balls, while multi-camera setups have facilitated precise 3D trajectory reconstruction. These technologies have dramatically improved the accuracy of ball tracking, speed analysis, and serve legality detection. However, current solutions are often constrained to basic data analysis and fail to provide deeper semantic insights or actionable guidance for players and coaches.

The integration of Large Language Models (LLMs) with CV technologies offers a transformative opportunity. LLMs, with their advanced capabilities in natural language understanding and generation, can interpret complex data outputs and translate them into meaningful insights or coaching advice. Leveraging detailed data on ball trajectories, player movements, and hitting techniques, such systems can help athletes and coaches pinpoint strengths, address weaknesses, and tailor strategies to specific match scenarios. For instance, personalized insights can guide players to refine techniques, enhance footwork, or adapt strategies against an opponent’s play style. These integrated systems not only boost training efficiency but also provide real-time decision-making support during matches.

This study is driven by the need to develop an integrated system that combines CV, deep learning, and LLMs to deliver comprehensive solutions for table tennis training, competition analysis, and officiating. By advancing AI-driven sports analytics, this research aims to establish a new standard for how data-driven tools support athletes and coaches in high-speed, high-precision sports.

## Research Questions

This study addresses critical challenges in table tennis analytics and aims to explore innovative solutions through the integration of computer vision (CV), deep learning, and large language models (LLMs). The primary research questions guiding this study are:

1. How can multi-camera systems enhance the precision of ball tracking and 3D trajectory reconstruction in table tennis?

This question seeks to evaluate the effectiveness of using synchronized multi-camera setups to overcome limitations such as occlusions and depth perception issues that hinder single-camera systems.

1. What are the key trajectory features and events (e.g., throw point, highest point, and hit point) that can be automatically identified to improve serve legality detection?

This question focuses on understanding the dynamics of the ball's motion and exploring how deep learning models, such as YOLO and Transformer, can detect critical trajectory points.

1. Can Large Language Models (LLMs) provide actionable insights and real-time feedback by interpreting raw data from computer vision models?

This question aims to determine how LLMs, fine-tuned with domain-specific data, can translate numerical and visual data into meaningful coaching advice, tactical suggestions, and rule compliance evaluation.

1. How can an integrated framework combining CV, deep learning, and LLMs support real-time decision-making during matches and training?

This question investigates the feasibility of creating an interactive system that not only analyzes data but also provides real-time feedback to athletes and coaches in high-speed, high-stakes scenarios.

By addressing these questions, this study aims to bridge the gap between advanced data analysis and practical applications in table tennis, ultimately transforming how athletes and coaches utilize technology to enhance performance and ensure fair play.

## Contributions

The This study presents several key contributions that advance the integration of computer vision, deep learning, and large language models (LLMs) in the context of table tennis analytics:

1. Efficient 3D Reconstruction for Dynamic Environments

A novel calibration method leveraging the inherent dimensions of the table tennis court was developed, enabling rapid 3D reconstruction of the playing environment during venue changes. This approach facilitates accurate tracking of player speed and activity range.

1. Accurate Player Action Identification and Classification

By applying Magnus effect principles and analyzing the 3D trajectory of the ball, this study provides a robust methodology for determining precise player actions and stroke types, significantly enhancing the understanding of player techniques.

1. YOLO Optimization for Small Object Scenarios

YOLO was optimized specifically for small, fast-moving objects like table tennis balls, resulting in improved accuracy in 3D tracking and detection. This enhancement addresses challenges unique to high-speed sports analytics.

1. Precise Video Segmentation Using 3D Trajectories

A trajectory-based approach was implemented to achieve fine-grained video segmentation, enabling action segmentation for table tennis players. This allows detailed analysis of specific movements and sequences during gameplay.

1. Transformer-Based Trajectory Analysis for Key Turning Points

Transformer models were applied to 3D trajectory data to identify critical turning points, such as the throw point, highest point, and hit point. This analysis supports accurate serve legality assessments and data collection for foul detection.

1. Development of a Domain-Specific Q/A Dataset

A custom dataset was constructed to train LLMs, incorporating match statistics, training suggestions, and strategic advice tailored to table tennis. This dataset ensures domain relevance and enhances the model’s performance.

1. LoRA Fine-Tuning for Table Tennis Scenarios

LoRA fine-tuning was employed to optimize the LLM for understanding and generating outputs specific to table tennis. This validated the feasibility of combining LLMs with computer vision for real-time analysis.

1. Interactive Intelligent Assistant for Match Data

An innovative application was developed to transform traditional data analysis tools into an interactive intelligent assistant. This system improves the interpretability and usability of match data, providing actionable insights and real-time guidance for players and coaches.

## Objectives of This Report

This report aims to tackle key challenges in table tennis analytics by proposing a comprehensive framework that integrates computer vision (CV), deep learning, and large language models (LLMs). A central objective is to design and implement a robust multi-camera system capable of precise 3D trajectory reconstruction and tracking of table tennis balls and players, even in high-speed scenarios. Such a system is intended to overcome limitations like occlusions and depth perception challenges inherent in traditional single-camera setups.

Another goal is to explore the application of advanced deep learning models, such as YOLO and Transformers, to improve small object detection and analyze ball trajectories. Special attention is given to identifying critical turning points, such as the throw, highest, and hit points, and leveraging these insights for accurate serve legality assessments. In parallel, the study seeks to develop methods for trajectory-based video segmentation, enabling detailed analysis of player actions and technical movements throughout a match.

In addition, the report focuses on integrating LLMs fine-tuned with domain-specific datasets to interpret the outputs of CV models and provide actionable feedback. This includes transforming raw data into meaningful insights, such as tactical recommendations, training plans, and real-time serve compliance evaluations. By combining these components, the ultimate objective is to create an interactive and intelligent system that supports players, coaches, and referees by delivering real-time decision-making assistance and enhancing the usability of AI-driven tools in table tennis.

## Structure of This Report

The description of this report is as follows:

* Chapter 2: Literature Review examines prior research in computer vision, deep learning, and large language models as applied to sports analytics. It highlights the limitations of existing methods and identifies the research gap this study aims to address.
* Chapter 3: Methodology details the experimental setup, including the multi-camera system, data collection processes, and calibration techniques for 3D trajectory reconstruction. This chapter also describes the optimization of YOLO and Transformer models, the construction of domain-specific datasets, and the fine-tuning of LLMs with LoRA.
* Chapter 4: Results presents the outcomes of the proposed framework, including metrics for ball detection and tracking accuracy, key trajectory point identification, serve compliance evaluation, and real-time interaction capabilities. It includes both quantitative results and qualitative insights gained through practical application.
* Chapter 5: Discussion analyzes the implications of the findings, comparing them with prior research. It explores the strengths and limitations of the proposed approach and discusses potential improvements for future studies.
* Chapter 6: Conclusion summarizes the key contributions and achievements of the study, emphasizing its impact on table tennis analytics. It also outlines possible directions for extending this research to other fast-paced individual sports.

# Chapter 2 Literature Review

*The focus of this report is on … based on .., this chapter will introduce a plenty of traditional methods and the relevant knowledge of ...*

## Introduction

As a cutting-edge research

.

## Xxxx1

Human ddd2…

## Xxxx2

Human dddd..

.

In summary,….

# Chapter 3 Methodology

*The main content of this chapter is to clearly articulate research methods, which satisfy the objectives of this report. The chapter mainly covers the details of research methodology for …. which will be clearly introduced with the confident and imaginative use of the feature description methods.*

## Xxxx1

In recognizing and ers.

## Xxxx1

rch.

## Model

ions.

## Evaluation

In order to provide

e posture.

## Dataset

A billiard

testing.

# Chapter 4 Results

*The main content of this chapter is to collect video data and demonstrate the experimental results. In the end, in this chapter, we also discuss the limitations of this project.*

## Xxxx1

This chapter

ture through comparative observation.

## Xxxx2

In the scenario.

|  |
| --- |
| Fig 4.4 Scoring system based on joint angle calculation. |

Figure 4.4 shows ke.

## Xxxx3

We have described the structure of Transformer model in the methodology section, next, we will discuss the

ns.

Table 4.1 Results of activation function ablation experiments.

|  |  |  |  |
| --- | --- | --- | --- |
| Activation Function | Accuracy | Loss | F1-Score |
| None | 98.32% | 0.0957 | 0.9831 |
| Mish | 89.72% | 0.3588 | 0.8971 |
| Swish | 90.28% | 0.4117 | 0.9028 |

From the results in Table 4.1, the model .

# Chapter 5 Analysis and Discussions

*In this chapter, experimental results are analyzed and compared. Comparisons of the results under various conditions will be explored.*

### Analysis

In summary, we combined ….

### Discussions

In our experiments。。。

d analysis.

# Chapter 6 Conclusion and Future Work

*In this chapter, we will summarize the subject and method of this project and propose new research direction according to the result and insufficiency of the experiment as well as the future work.*

## Conclusion

This report aims to combine human pose estimation and realize the striking pose and analysis of a billiard player onds.

## Future Work

We will increase.

# References

Angelini, F., Fu, Z., Long, Y., Shao, L., & Naqvi, S. M. (2019). 2D pose-based real-time human action recognition with occlusion-handling. *IEEE Transactions on Multimedia*, 22(6), 1433-1446.

Abujar, S., Masum, A. K. M., Chowdhury, S. M. H., Hasan, M., & Hossain, S. A. (2019). Bengali text generation using bi-directional RNN. In International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5).

Agrawal, S., & Sharma, D. K. (2022). Feature extraction and selection techniques for time series data classification: A comparative analysis. In International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 860-865).

Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2D human pose estimation: New benchmark and state of the art analysis. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 3686-3693).

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166.

Badiola-Bengoa, A., & Mendez-Zorrilla, A. (2021). A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. Sensors, 21(18), 5996.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877-1901.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:2006.10204.

Borodulina, A. (2019). Application of 3D Human Pose Estimation for Motion Capture and Character Animation (Master's Thesis) University of Oulu, Finland.

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2019). Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. Journal of Sports Sciences, 37(5), 568-600.

Cao, X. (2022) Pose Estimation of Swimmers from Digital Images Using Deep Learning. *Master’s Thesis, Auckland University of Technology*.

Cao, X., and Yan, W. (2022) Pose estimation for swimmers in video surveillance. *Multimedia Tools and Applications, Springer*.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *In IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7291-7299).

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Cooijmans, T., Ballas, N., Laurent, C., Gülçehre, Ç., & Courville, A. (2016). Recurrent batch normalization. *arXiv preprint arXiv:1603.09025*.

Dang, Q., Yin, J., Wang, B., & Zheng, W. (2019). Deep learning based 2d human pose estimation: A survey. Tsinghua Science and Technology, 24(6), 663-676.

Dalal, N., & Triggs, B. (2005,). Histograms of oriented gradients for human detection. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 886-893).

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. *In IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2625-2634).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Ekvall, S., Kragic, D., & Hoffmann, F. (2005). Object recognition and pose estimation using color cooccurrence histograms and geometric modeling. Image and Vision Computing, 23(11), 943-955.

Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14(2), 179-211.

Fragkiadaki, K., Levine, S., Felsen, P., & Malik, J. (2015). Recurrent network models for human dynamics. *In IEEE International Conference on Computer Vision* (pp. 4346-4354).

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural Computation, 12(10), 2451-2471.

Gholamalinezhad, H., & Khosravi, H. (2020). Pooling methods in deep neural networks, a review. arXiv preprint arXiv:2009.07485.

Gowdra, N. (2021) Entropy-Based Optimization Strategies for Convolutional Neural Networks. PhD Thesis, Auckland University of Technology, New Zealand.

Graves, A., & Graves, A. (2012). Long short-term memory. In Supervised Sequence Labelling with Recurrent Neural Networks, 37-45.

Gu, D., Nguyen, M., Yan, W. (2016) Cross models for twin recognition. International Journal of Digital Crime and Forensics 8 (4), 26-36.

Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *In IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645-6649).

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6), 602-610.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. Diploma, Technische Universität München, 91(1), 31.

Herrera, A., Beck, A., Bell, D., Miller, P., Wu, Q., Yan, W. (2008) Behavior analysis and prediction in image sequences using rough sets. International Machine Vision and Image Processing Conference (pp.71-76)

Hori, T., Cho, J., & Watanabe, S. (2018). End-to-end speech recognition with word-based RNN language models. In IEEE Spoken Language Technology Workshop (SLT) (pp. 389-396).

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *In IEEE International Conference on Computer Vision* (pp. 2961-2969).

Illavarason, P., Arokia Renjit, J., & Mohan Kumar, P. (2019). Medical diagnosis of cerebral palsy rehabilitation using eye images in machine learning techniques. *Journal of Medical Systems*, *43*(8), 278.

Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). An empirical exploration of recurrent network architectures. *In International Conference on Machine Learning* (pp. 2342-2350). PMLR.

Jian, S., Kaiming, H., Shaoqing, R., & Xiangyu, Z. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision & Pattern Recognition* (pp. 770-778).

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Liu, T., Wang, K., Sha, L., Chang, B., & Sui, Z. (2018, April). Table-to-text generation by structure-aware seq2seq learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012-10022).

Lee, K., Kim, W., & Lee, S. (2022). From human pose similarity metric to 3D human pose estimator: Temporal propagating LSTM networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2), 1781-1797.

Lin, Y., Jiao, X., & Zhao, L. (2023). Detection of 3D human posture based on improved MediaPipe. Journal of Computer and Communications, 11(2), 102-121.

Li, F., Zhang, Y., Yan, W., Klette, R. (2016) Adaptive and compressive target tracking based on feature point matching. International Conference on Pattern Recognition (ICPR), (pp.2734-2739).

Liu, C., Yan, W. (2020) Gait recognition using deep learning. Handbook of Research on Multimedia Cyber Security (pp.214-226)

Liu, Y., Nand, P., Hossain, A., Nguyen, M., Yan, W. (2023) Sign language recognition from digital videos using feature pyramid network with detection Transformer. *Multimedia Tools and Applications*.

Lu, J. (2016) *Empirical Approaches for Human Behavior Analytics*. Master’s Thesis. Auckland University of Technology, New Zealand.

Lu, J., Shen, J., Yan, W., Boris, B. (2017) An empirical study for human behaviors analysis. *International Journal of Digital Crime and Forensics*, 9 (3), 11-17.

Lu, J., Nguyen, M., Yan, W. (2018) Pedestrian detection using deep learning. IEEE AVSS.

Lu, J., Nguyen, M., Yan, W. (2020) Human behavior recognition using deep learning. International Conference on Image and Vision Computing New Zealand.

Lu, J., Nguyen, M., Yan, W. (2020) Comparative evaluations of human behavior recognition using deep learning. *Handbook of Research on Multimedia Cyber Security*, 176-189.

Lu, J., Nguyen, M., Yan, W. (2021) Sign language recognition from digital videos using deep learning methods. *International Symposium on Geometry and Vision.*

Lu, J. (2021) *Deep Learning Methods for Human Behavior Recognition.* PhD Thesis. Auckland University of Technology, New Zealand.

Mao, W., Ge, Y., Shen, C., Tian, Z., Wang, X., & Wang, Z. (2021). TFPose: Direct human pose estimation with transformers. *arXiv preprint arXiv:2103.15320*.

Neuman, B., & Gray, R. (2013). A direct comparison of the effects of imagery and action observation on hitting performance. Movement & Sport Sciences-Science & Motricité, (79), 11-21.

Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. *In European Conference on Computer Vision* (pp. 483-499). Springer International Publishing.

Nie, B. X., Wei, P., & Zhu, S. C. (2017). Monocular 3d human pose estimation by predicting depth on joints. *In IEEE International Conference on Computer Vision (ICCV)* (pp. 3467-3475).

Özgür, A., & Nar, F. (2020). Effect of dropout layer on classical regression problems. In Signal Processing and Communications Applications Conference (SIU) (pp. 1-4).

Parmar, P., & Tran Morris, B. (2017). Learning to score Olympic events. *In IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 20-28).

Pfeiffer, M., Paolo, G., Sommer, H., Nieto, J., Siegwart, R., & Cadena, C. (2018). A data-driven model for interaction-aware pedestrian motion prediction in object cluttered environments. *In IEEE International Conference on Robotics and Automation (ICRA)* (pp. 5921-5928).

Pan, C., Yan, W. (2018) A learning-based positive feedback in salient object detection. International Conference on Image and Vision Computing New Zealand.

Pan, C., Yan, W. (2020) Object detection based on saturation of visual perception. Multimedia Tools and Applications, 79 (27-28), 19925-19944.

Pan, C., Liu, J., Yan, W., Zhou, Y. (2021) Salient object detection based on visual perceptual saturation and two-stream hybrid networks. IEEE Transactions on Image Processing.

Ramanan, D., Forsyth, D. A., & Zisserman, A. (2006). Tracking people by learning their appearance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1), 65-81.

Rossi, A., Pappalardo, L., & Cintia, P. (2021). A narrative review for a machine learning application in sports: an example based on injury forecasting in soccer. Sports, 10(1), 5.

Siddiqui, H. U. R., Younas, F., Rustam, F., Flores, E. S., Ballester, J. B., Diez, I. D. L. T., ... & Ashraf, I. (2023). Enhancing cricket performance analysis with human pose estimation and machine learning. Sensors, 23(15), 6839.

Sun, X., Wu, P., & Hoi, S. C. (2018). Face detection using deep learning: An improved Faster R-CNN approach. Neurocomputing, 299, 42-50.

Song, S., Lan, C., Xing, J., Zeng, W., & Liu, J. (2017). An end-to-end spatio-temporal attention model for human action recognition from skeleton data. *In AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27.

Shen, J., Yan, W., Miller, P., Zhou, H. (2010) Human localization in a cluttered space using multiple cameras. IEEE International Conference on Advanced Video and Signal Based Surveillance.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Tang, B., & Matteson, D. S. (2021). Probabilistic transformer for time series analysis. Advances in Neural Information Processing Systems, 34, 23592-23608.

Toshev, A., & Szegedy, C. (2014). DeepPose: Human pose estimation via deep neural networks. *In IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1653-1660).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., & Shao, L. (2021). Deep 3D human pose estimation: A review. Computer Vision and Image Understanding, 210, 103225.

Wu, Q., Xu, G., Zhang, S., Li, Y., & Wei, F. (2020). Human 3D pose estimation in a lying position by RGB-D images for medical diagnosis and rehabilitation. In Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 5802-5805).

Wang, X., Yan, W. (2019) Human gait recognition based on self-adaptive hidden Markov model. IEEE/ACM Transactions on Biology and Bioinformatics.

Wang, X., Yan, W. (2019) Cross-view gait recognition through ensemble learning. Neural Computing and Applications.

Wang, X., Yan, W. (2019) Gait recognition using multichannel convolutional neural networks. Neural Computing and Applications.

Wang, X., Yan, W. (2019) Multi-perspective gait recognition based on ensemble learning. Springer Neural Computing and Applications.

Wang, X., Yan, W. (2019) Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory. International Journal of Neural Systems.

Wang, X., Yan, W. (2020) Non-local gait feature extraction and human identification. Springer Multimedia Tools and Applications.

Wang, X., Yan, W. (2020) Cross-view gait recognition through ensemble learning. Neural Computing and Applications 32 (11), 7275-7287.

Wang, X., Yan, W. (2022) Human identification based on gait manifold. Applied Intelligence.

Wöllmer, M., Kaiser, M., Eyben, F., Schuller, B., & Rigoll, G. (2013). LSTM-modeling of continuous emotions in an audiovisual affect recognition framework. *Image and Vision Computing*, 31(2), 153-163.

Xiao, B., Wu, H., & Wei, Y. (2018). Simple baselines for human pose estimation and tracking. *In European Conference on Computer Vision (ECCV)* (pp. 466-481).

Yan, W. (2023) Computational Methods for Deep Learning. Springer.

Yan, W. (2019) Introduction to Intelligent Surveillance. Springer.

Yu, Q., & Guo, H. (2022). Sports medicine image modeling for injury prevention in basketball training. Contrast Media & Molecular Imaging, 2022.

Yu, Z. (2021) Deep Learning Methods for Human Action Recognition. Master’s Thesis, Auckland University of Technology, New Zealand.

Yu, Z., Yan, W. (2020) Human action recognition using deep learning methods. International Conference on Image and Vision Computing New Zealand.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural Computation, 31(7), 1235-1270.

Zhi-chao, C., & Zhang, L. (2019). Key pose recognition toward sports scene using deeply-learned model. *Journal of Visual Communication and Image Representation*, 63, 102571.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *In ECCV* (pp. 818-833).

Zhu, Y., Yan, W. (2022) Ski fall detection from digital images using deep learning. ACM ICCCV.

Zheng, C., Zhu, S., Mendieta, M., Yang, T., Chen, C., & Ding, Z. (2021). 3D human pose estimation with spatial and temporal transformers. In*Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 11656-11665).