What pose estimation methods are most effective for analysing cricket shots?

Daniel Plevier
EEMCS, Delft University of Technology

Responsible Professor:Ujwal Gadiraju, Supervisor: Danning Zhan

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Daniel Plevier
Final project course: CSE3000 Research Project
Thesis committee: Ujwal Gadiraju, Danning Zahn, Marx Neerincx

An electronic version of this thesis is available at http://repository.tudelft.nl/.

EEMCS, Delft University of Technology, The Netherlands.

Abstract

Classification of cricket shots is a recent field of study that has seen some growth. The addition of Human Pose Estimation (HPE) has the potential to advance the study of cricket shot classification. This paper investigates which of the available and widely use HPE frameworks are most suitable for for the domain of cricket and introduces a hand annotated verification dataset in order to test this. The findings indicate that any increase in precision will come at a steep computational cost but a middle ground can be found depending on the use-case.

1 Introduction

Pose estimation and motion capture have a long history in the field of computer science. Research in this area is still developing at a rapid pace with many new methods being proposed in a wide variety of fields such as sports and (sports)broadcasting, rehabilitation and animation. A large number of trade-off's are however still required in most human pose estimation methods, this can often result in a high computational cost.[10] For the purpose of sports analysis pose estimation is already used to analyze the performance of athletes and provide insights for further improvements. These human pose estimation methods can already achieve a high degree of precision in classifying different stroke techinques used.[11] Some of the most recent study is also devoted to Pose estimation under less than ideal circumstances closer to real world situations. One of the examples is data that is partly occluded. These slightly covered images require inference of joints.[4] Other fields of study pertain to enhancing 3D pose estimation methods by combining multiple views or performing pose estimation on scenes with multiple people in them [5]

In the field of cricket a common task for visual images is the classification of shots for the purpose of training and analysis. [7] In order to enhance these classifications recent papers have proposed adding pose estimation in order to enhance the accuracy.[2] [11] None of these papers however motivate their choice of pose estimation method beyond the fact that it is a well known pose estimation method.

This study will investigate the question **what pose estimation methods are most effective for analyzing cricket shots?** This question will be answered through splitting it up into 3 smaller sub-questions.

1. What existing frameworks or studies have explored pose estimation in sports or other dynamic domains?
2. What are the computational and practical trade-offs (e.g., accuracy, latency) between different pose estimation methods?
3. How well do different pose estimation techniques handle real-world challenges, such as occlusions and variations in camera angles?

In section 2 we wil delve into the top of the line methods of pose estimation associated datasets and real life problems that with them also will the method for our research be explained. In section 3 the general motivation and structure of our experiment will be explained. Section 4 contains an in dept description of the process while section 5 will discuss the results 6 talk about this papers commitment to responsible research and section 7 will contain the final discussion and possible avenue's for future research

2 Background

2.1 pose estimation models

For this paper 3 popular HPE frameworks will be compared. Mediapipe, Openpose and HRNet are 3 HPE frameworks that are often used in dyanamic domains and sports related apllications. (cite) HRNet and its some of the papers that build on it are consistently among the best performing methods in papers [6] that focos on general datasets like OCHuman and COCO[8]

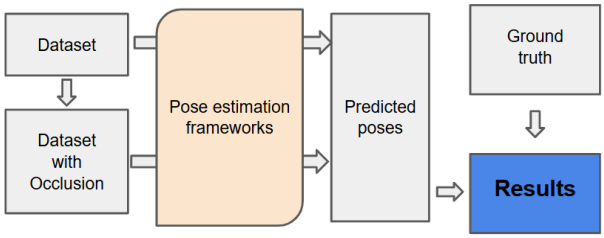


Figure 1: research pipeline

2.2 datasets

Several datasets exist which can be used to train and validate human pose estimation (HPE) frameworks. The One thing of note is however that none of these contain any amount of training data for cricket. Most papers that use HPE and cricket in conjunction try to classify the shot of a cricket player. Because of this there are some small datasets that classify cricket shots and these use HPE to enhance the classification but none of them test the effectiveness of HPE for the domain of cricket. Because of this we will create a hand anotated-dataset that has the keypoints for several hundred cricket players.

2.2.1 COCO

COCO[8] is a very large annotated dataset that contains a large amount of hand-annotated images. The images contained range over a broad variety of topics and are suitable for a variety of computer vision tasks such as object detection, pose estimation, classification and captioning.

2.3 real world challenges in human pose estimation

One of the aspect of HPE that is bein actively researched is occlusion.[13] As opposed to low light conditions or highly crowded spaces this is quite relevant to cricket where another player might cover part of the human that is being read or the person self might be in a position where some part of their body covers another part. The ability to compensate for that is a large part of what would make a HPE system usefull for cricket analasys

3 Method

In order to evaluate the chosen we follow the pipeline as show in figure 1 1. First we take our original dataset. Part of the players in this dataset are occluded on the joints in order to simulate an occulded dataset. Then both the normal and occluded datasets are run through the HPE frameworks in order to recieve a predicted pose in all situations. These results are then compared to the recorded ground truths for these images. The general perforance of the frameworks is also measured. From these the results are collected.

3.1 pose estimation models

HRNet[12] consistently ranks among the best in any survey about human pose estimation[1][6] which is why it will be chosen for the purpose of this paper. Mediapipe and Openpose are selected because they have been used in other cricket related research papers.[9][11]

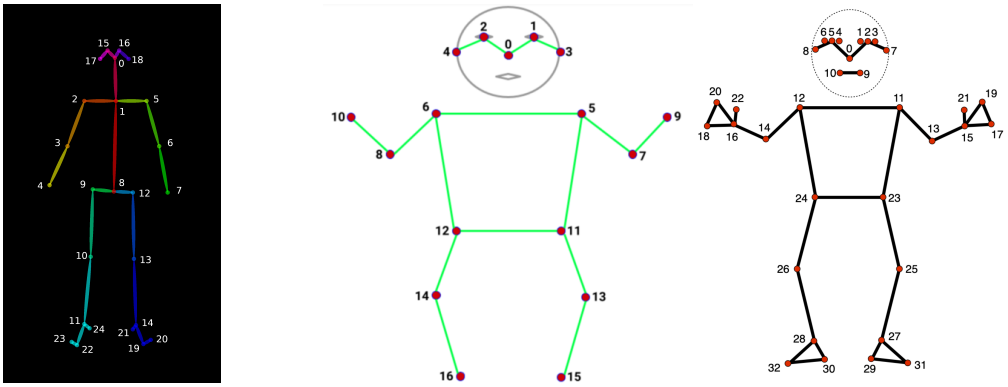
3.2 datasets

Within the literature there is a clear trend for datasets. The large datasets such as COCO[] have a very varied composition and are useful for the training of large models. The COCO dataset does however not contain any instances of cricket despite the large representation of some other sports such as soccer and basketball as well as some baseball. Other large datasets are equally unsuitable for cricket specific pose estimation.

In the field of cricket shot classification there do exist databases but these do not include a ground truth for HPE. In the few papers that make use of pose estimation in their classification process they cannot test the resulting keypoints against a ground truth.

As part of this research paper a small hand annotated dataset will be made containing only cricket players. Using this we can better determine if a given HPE framework is truly suitable for tasks involving cricket. In order to test the resilience of these frameworks a secondary dataset wil also be made by occluding part of the first dataset similar to "Limb joint augmentation"[4].

3.3 keypoints



(a) fig:Keypoints from Openpose

(b) Standard COCO keypoints also used by HRNet

(c) Keypoints from Mediapipe

Figure 2: The different keypoint skeletons estimated by the different HPE's

For the purpose of the comparison all of the keypoint skeletons are converted to the COCO skeleton. Since this is a subset of the other 2 and COCO a common standard is this wil cause the least problems.

4 Experiment

All the following experiments were performed on a personal computer with a Intel(R) Core(TM) i7-12700H CPU and a GeForce RTX™ 3050Ti Laptop GPU with 16 gb of RAM.

4.1 creating dataset

1. First, a data set of images is collected from kaggle [1]. This is a classification data set of 4724 images in 4 classes that do not have keypoint annotation but are suitable for our purposes since they mainly have single person pictures without people in the background.
2. From these images a subset is taken of 191 pictures from all 4 classes mostly by filtering out augmented duplicates.
3. Using cvat.ai[2] a ground truth keypoints are created for all the 191 images in the dataset. These are then exported in the COCO standard and form the basis of our self made dataset.

4.1.1 occluding dataset

1. In every image a random keypoint is selected (head counts as 1 keypoint considering that it consists of several closery clustered together)
2. A black sphere the size of 15% of the image's height is projected in order to occlude the image below as seen in Figure 3

[1] https://www.kaggle.com/datasets/aneesh10/cricket-shot-dataset
[2] https://www.cvat.ai/

(a) regular image

(b) occluded image

Figure 3: Example of the occlusion in the dataset

4.2 running human pose estimation

4.2.1 HRNet

Simple HRNet is used for it's greater ease of use and availability of supporting documentation. The x and y co-ordinates are swapped compared to the other PHE Frameworks but since the output is already in the COCO_17 format it is a simple matter to fix.

Mediapipe

using the notebook from the documentation the network can be started quite easily. For some unknown reason Mediapipe seems to be unable to process certain images from the dataset. Simply returning an empty list. This is 30 in the regular dataset and 78 in the occluded dataset. It is unknown why this is.

4.2.2 OpenPose

4.3 resulting metrics

In order to compare the measured results from our pose estimation frameworks to the annotated ground truth we use object keypoint simalarity. This is one of the most prominent ways to measure simalarity in the field of HPE[3]

$$\text{OKS} = \frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right) \cdot v_i}{\sum_i v_i}$$

The time that each model runs is recorded and saved as the number of seconds that it runs. Since all pose estimation models use the same dataset this will give a good impression of the overall computational cost.

5 Results

	Runtime (s)	Average OKS
MediaPipe	31.4	0.50
HRNet	420.9	0.65
OpenPose	15.8	

5.1 obscured

	Runtime (s)	Average OKS
MediaPipe	29.1	0.25
HRNet	393.6	0.55
OpenPose	16.0	

6 Responsible Research

In order to improve the reproducibility of the paper all notebooks and other code snippets that where used during this project are shared on the authors Github.[3] The dataset that was newly created for the purpose of this project is also available here. The frameworks and tools that where utilized during this project are all referenced where they are mentioned.

This paper was made with the purpose of benefiting any project that makes use of human pose estimation especially within the domain of cricket. While great care was taken to avoid this there might still be errors in the assumptions that are made during the process, and subsequently the conclussions might still be wrong.

7 Discussion

Although HRNet has the best results out of the tested HPE frameworks. The significantly longer runtime does seriously limit the applications that it is suitable for. A single image to annotate might be fine but is will be impossible to keep up with a live video feed.

7.1 future research

Based on the research performed in this paper there are still several directions that it can expand into. For now they are outside the scope of this paper but further work might be warranted to enhance the results.

7.1.1 Real world application

Pose estimation is seldom the intended purpose of any application. Further research could be done in how the methods that are recommended here perform when they feed into other research such as classification or providing a visual aid during training. This might still reveal issues that where not taken into account in this paper.

7.1.2 General expansion of scope

Due to limitations this paper evaluated only a limited number of HPE frameworks, created only a minor dataset to support these evalations and tested only the effect of occlusion on the performance of these datasets. More HPE frameworks, larger datasets and a wider variety of real world challenges could still be valuable in general.

7.1.3 Kneeling and padding

During the process of evaluating the results of the HPE it was noticed that on several occasions the frameworks struggled to correctly estimate limbs that where partially obscured by them self for example kneeling. The fact that cricket players usually also wear heavy padding seems to cause issues for the HPE frameworks. It might be worth it to study how much performance is lost because of this.

References

[1] Haoming Chen, Runyang Feng, Sifan Wu, Hao Xu, Fengcheng Zhou, and Zhenguang Liu. 2d human pose estimation: A survey. *Multimedia systems*, 29(5):3115–3138, 2023.

[2] Ks Sreekar Datta, G Narasimha Naidu, Rahul Dasari, and T V Jayakumar. Advancements in Cricket Shot Detection: Integrating Human Pose Estimation and Deep Learning for Automated Analysis. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–7, June 2024. ISSN: 2473-7674.

[3] https://github.com/DanielPlevier/Cricket$_{p}ose_{e}stimation$

[3] Kerui Gu, Rongyu Chen, and Angela Yao. On the calibration of human pose estimation. *arXiv preprint arXiv:2311.17105*, 2023.

[4] Gangtao Han, Chunxiao Song, Song Wang, Hao Wang, Enqing Chen, and Guanghui Wang. Occluded human pose estimation based on limb joint augmentation. *Neural Computing and Applications*, 37(3):1241–1253, January 2025.

[5] Taemin Hwang, Jieun Kim, Myoungjin Kim, and Minjoon Kim. A Real-time Multi-Person 3D Pose Estimation System from Multiple RGB-D Views for Live Streaming of 3D Animation. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23 Companion, pages 105–107, New York, NY, USA, 2023. Association for Computing Machinery.

[6] Pawel Knap. Human Modelling and Pose Estimation Overview, June 2024. arXiv:2406.19290 [cs].

[7] Ashok Kumar, Javesh Garg, and Amitabha Mukerjee. Cricket activity detection. In *International Image Processing, Applications and Systems Conference*, pages 1–6, November 2014.

[8] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common Objects in Context, February 2015. arXiv:1405.0312 [cs].

[9] Tevin Moodley and Dustin van der Haar. Casrm: cricket automation and stroke recognition model using openpose. In *International Conference on Human-Computer Interaction*, pages 67–78. Springer, 2020.

[10] Ana Filipa Rodrigues Nogueira, Hélder P. Oliveira, and Luís F. Teixeira. Markerless multi-view 3D human pose estimation: A survey. *Image and Vision Computing*, 155:105437, March 2025.

[11] Hafeez Ur Rehman Siddiqui, Faizan Younas, Furqan Rustam, Emmanuel Soriano Flores, Julién Brito Ballester, Isabel de la Torre Diez, Sandra Dudley, and Imran Ashraf. Enhancing Cricket Performance Analysis with Human Pose Estimation and Machine Learning. *Sensors*, 23(15):6839, January 2023. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

[12] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5693–5703, 2019.

[13] Song-Hai Zhang, Ruilong Li, Xin Dong, Paul L. Rosin, Zixi Cai, Xi Han, Dingcheng Yang, Hao-Zhi Huang, and Shi-Min Hu. Pose2Seg: Detection Free Human Instance Segmentation, April 2019. arXiv:1803.10683 [cs].