A computer-vision based training coach for computerized physical training

George Davies[0009−0008−4132−5676]

University of Lincoln, UK 27421138@students.lincoln.ac.uk

1 Introduction

In many stories of science-fiction, there are robots that are intelligent, borderline human in the way they talk, the way they walk, and the way they interact with their environments. These kinds of robots seem to be creatures that won't exist any time soon, and they probably won't, but with every advancement in the field of robotics and autonomous systems, we step closer to that reality. One key sense that helps a robot communicate and understand its environment is sight, and the area of robotics that strives towards the goal of giving machines the sense of sight is computer vision.

1.1 Area of Research

Human Pose Estimation (HPE) is an area of research within computer vision that aims to teach robots how to make sense of the human form and the motions it is capable of performing. It involves the identification and classification of the joints in the human body, capturing a set of coordinates for each joint, known as a key point, that can describe the pose of a person. HPE has a wide set of uses in many fields: In games, with motion capture technologies reliant on HPE, it allows developers to code program more realistic and fluid character movements. In healthcare, healthcare providers can monitor a patient's movements and detect any abnormalities. Augmented reality, allows the user to interact with the digital content in more natural and intuitive ways with gestures. And finally, the use-case that is the primary focus of this project, is sports training. HPE can be used to analyse a user's performance, identify areas for improvement, and develop personalised training programs based on the physical level of the user. For example, HPE could be used to analyse a runner's form, e.g. How straight is their back? What part of the foot are they landing on? Are they leaning more to one side?..., and provide feedback on how to improve their technique. HPE can be used to collect data about any exercises where the movement of the body is vital to its effectiveness.

1.2 Relevance to the Degree Programme

I am enrolled in the MSc programme Robotics and Autonomous Systems at the Univerity of Lincoln through the AgriFoRwArdS CDT. Throughout the first

two semesters, I studied the principles of robotics, artificial intelligence, machine learning, and computer vision. Principles from all of these subjects are applied within the area of research on human pose estimation. As to my affiliation with the AgriFoRwArdS CDT, their focus is on the production and use of AI, ML, and CV applications to help the farming and agricultural industries, as Lincolnshire is an agricultural region of the UK. Human pose estimation has previously been used in agritech applications (Moysiadis et al. 2022), with its ability to facilitate human-robot interactions in the field for fruit picking, robotic carts will follow the worker through the field to hold the produce and take it away once full. HPE allows these robots to understand gesture commands the worker may give it, and gives the robot an understanding of humans that allows it to find and follow them without driving into them.

1.3 Background of the Topic

When computer vision gained popularity in the late 1960s and early 1970s, HPE research had its start. Scientists first focused on basic problems like as shape analysis, object recognition, and visual understanding. As computer vision developed, HPE became a stand-alone area of study (Lynn 2023). Historically, HPE was frequently described probabilistically to account for likely inference ambiguities. Since deep learning has been more widely used, the focus has switched to end-to-end trainable models because of their ability to extract intricate patterns and postures from data. Traditionally, computer vision systems have assessed an object's or person's posture by geometric calculations and feature-based techniques. But, the biggest developments in HPE came with the advent of deep neural networks, convolutional neural networks, and computer vision. The field has advanced considerably in spite of these challenges, and more recent techniques that make use of properly designed neural networks may provide amazing results in challenging scenarios involving a large number of, perhaps veiled, interacting individuals (Liu and Yuan 2018). Now that these detections have the necessary technology and are sufficiently precise, they may be employed for commercial purposes. It also offers a wealth of new application potential and signifies a major change in HPE's overall direction.

2 Aims and Objectives

2.1 Issues to Explore

2.2 Motivation

2.3 End Goal

3 Literature Survey

3.1 BlazePose: On-device Real-time Body Pose tracking (Bazarevsky et al. 2020)

This 2020 Google Research by Valentin Bazarevsky et al. demonstrates a particular type of convolutional neural network architecture called BlazePose which

was created specifically for the real-time mobile human posture estimation task. With the ability to build 33 body keypoints for an individual and operating at over 30 frames per second on a Pixel 2, it is perfect for real-time applications such as fitness tracking and sign language recognition. BlazePose's primary contributions are a lightweight body pose estimate neural network and a novel body pose tracking method. Regression and heatmaps are both used by the network to determine keypoint coordinates, which probably improves efficiency and accuracy. An important development in the realm of on-device real-time body posture monitoring is this technology. It is a useful instrument for a variety of applications due to its efficiency and adaptability. BlazePose has a bright future ahead of it, and it will be interesting to watch how this technology develops and is applied in many real-world scenarios.

Fig. 1. BlazePose results on yoga and fitness poses

3.2 Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields (Cao et al. 2017)

In this 2017 paper by Zhe Cao et al., an effective technique for predicting posture for several people is Real-Time Multi-Person 2D posture estimate using Part Affinity Fields (PAFs) is presented. This method seeks to identify several individuals' 2D postures inside a picture. The system gains the ability to link body parts with specific persons in the image by using a nonparametric representation called Part Affinity Fields (PAFs). PAFs, or groups of 2D vector fields, represent the limbs' orientation and posture inside an image. Regardless of the number of persons in the image, the system's architecture for a greedy bottom-up parsing phase that produces exceptional accuracy and real-time speed.

It is therefore especially appropriate for use cases that include real-time data. The main contribution of this strategy is the separation of image population and runtime complexity. Compared to previous methods, which often saw increased computational costs as the number of participants increased, this represents a

Fig. 2. Top: Multi-person pose estimation. Body parts belonging to the same person are linked. Bottom left: Part Affinity Fields (PAFs) corresponding to the limb connecting right elbow and right wrist. The color encodes orientation. Bottom right: A zoomed in view of the predicted PAFs. At each pixel in the field, a 2D vector encodes the position and orientation of the limbs.

significant breakthrough. In summary, real-time multi-person 2D posture estimation using Part Affinity Fields presents this innovative and efficient approach for multi-person posture estimation. Because it can do real-time inference regardless of the amount of people in the image, it is a helpful tool in its sector.

3.3 Simple Baselines for Human Pose Estimation and Tracking (Xiao, Wu, and Wei 2018)

Simple Baselines for Human Pose Estimation and Tracking by Bin Xiao et al. is a significant addition to the field of human pose estimation. This study provides baseline methods that are practical and easy to apply for coming up with and evaluating new ideas in the sector. The main contribution of this study is the baseline tracking and posture estimation algorithms established. These methods are surprisingly effective, producing state-of-the-art performance on challenging benchmarks.

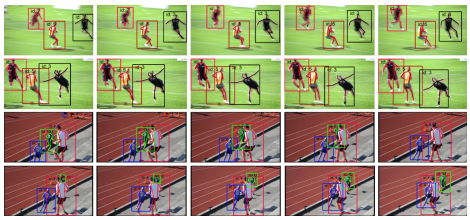

Fig. 3. Some sample results on PoseTrack Challenge test set

This approach's simplicity is one of its main advantages. This work shows that a straightforward approach may still provide outstanding results, even in the

face of the growing complexity of algorithms and systems in this sector. For the purpose of developing and assessing new techniques, this simplifies the process of analysing and comparing algorithms. It can be concluded that Simple Baselines for Human Pose Estimation and Tracking offers a productive and successful method for these tasks. Its ease of use and potency make it a useful instrument in its field, and there is a lot of room for expansion.

3.4 Pose2Seg: Detection Free Human Instance Segmentation (Zhang et al. 2019)

Pose2Seg, created by Song-Hai Zhang et al., is an innovative approach to human instance segmentation. This method provides a unique posture-based instance segmentation framework for humans, which divides instances based on human position, as opposed to proposal region identification. Usually, image instance segmentation begins with object detection and proceeds to segment the object from the detection bounding-box. Conversely, Pose2Seg takes into account the uniqueness of the "human" category, which is precisely specified by the posture skeleton. The human posture skeleton may be used to identify instances with severe occlusion instead of bounding-boxes.

Fig. 4. Heavily occluded people are better separated using human pose than using bounding-box.

This method shows that the pose-based framework can better handle occlusion and obtain higher accuracy on the human instance segmentation test compared to the most sophisticated detection-based approach. Additionally, the publication "Occluded Human (OCHuman)" introduces a new benchmark for occluded persons with full annotations, including instance masks, bounding boxes, and human position. This dataset has 8110 meticulously annotated human occurrences dispersed throughout 4731 images. All things considered, Pose2Seg offers a fresh and effective method for segmenting human instances. Its detection-free segmentation based on human postural capability makes it a top tool in its field with a lot of future application possibilities.

Summary The science of human posture evaluation has greatly benefited from these four publications. BlazePose makes real-time body location tracking of mobile devices easy and effective. To illustrate the effectiveness of nonparametric

representations in the precise estimation of multi-person poses, Realtime Multi-Person 2D Pose Estimation uses component affinity fields. The realistic baseline techniques for posture estimation and tracking provided by Simple Baselines for Human posture Estimation and Tracking emphasise the importance of simplicity in obtaining high performance. Pose2Seg offers a unique pose-based instance segmentation framework and is the last illustration of the potential for detection-free human instance segmentation. By building on the concepts of others, proposing fresh ideas, and refining previously established methods, each study advances the subject of human pose assessment.

4 Research Methods

5 Ethical Considerations

6 Project Plan and Risk Analysis

6.1 Project Plan

6.2 Risk Analysis

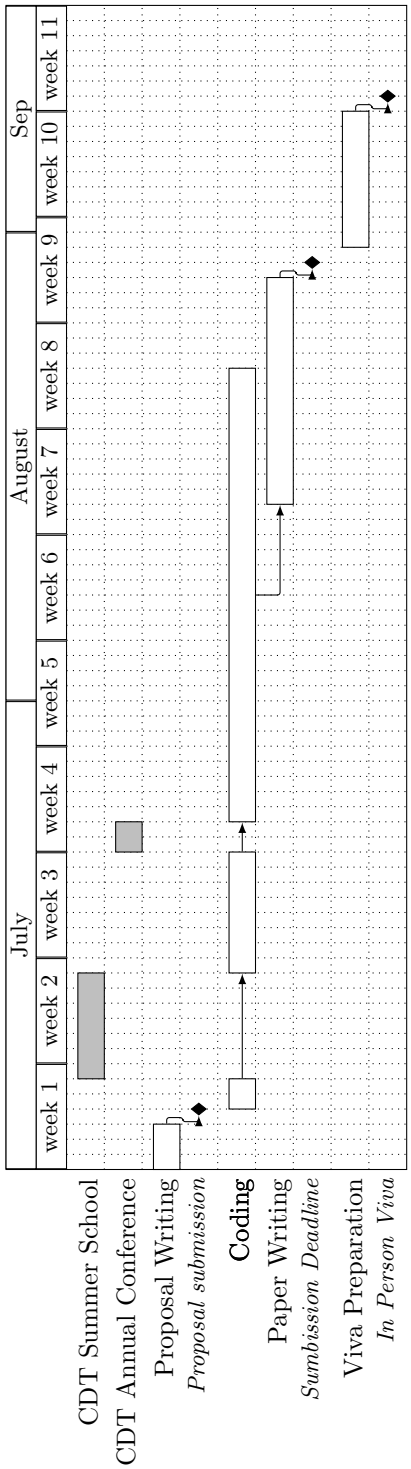

Fig. 5. Gantt Chart

References

Bazarevsky, Valentin et al. (2020). *BlazePose: On-device Real-time Body Pose tracking*. arXiv: 2006.10204 [cs.CV]. URL: https://arxiv.org/abs/2006.10204.

Cao, Zhe et al. (2017). *Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*. arXiv: 1611.08050 [cs.CV]. URL: https://arxiv.org/abs/1611.08050.

Liu, Mengyuan and Junsong Yuan (2018). "Recognizing human actions as the evolution of pose estimation maps". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1159–1168.

Lynn, Trevor (July 2023). *Pose Estimation Algorithms: History and Evolution.* Roboflow Blog. URL: https://blog.roboflow.com/pose-estimation-algorithms-history/.

Moysiadis, Vasileios et al. (2022). "An Integrated Real-Time Hand Gesture Recognition Framework for Human–Robot Interaction in Agriculture". In: *Applied Sciences* 12.16. ISSN: 2076-3417. DOI: 10.3390/app12168160. URL: https://www.mdpi.com/2076-3417/12/16/8160.

Xiao, Bin, Haiping Wu, and Yichen Wei (2018). *Simple Baselines for Human Pose Estimation and Tracking*. arXiv: 1804.06208 [cs.CV]. URL: https://arxiv.org/abs/1804.06208.

Zhang, Song-Hai et al. (2019). *Pose2Seg: Detection Free Human Instance Segmentation*. arXiv: 1803.10683 [cs.CV]. URL: https://arxiv.org/abs/1803.10683.

Word Count: 1730