PE Class Assistant
【AI Rating&Analysis】

Group 5

PE Class Assistant

AI Rating and Analysis

ygon7645 480481728

qgon2753 490539451

rlin3331 480084231

Table of Contents

1. BACKGROUND

Human pose estimation (recognition), is a set of processes to estimating the human body (pose) from images, videos in any background. It's actually a very important problem in computer visions. Lots of scientists were working to solve this and bring a better solution for computer visions for almost 15 years. This technology enables a higher level of human-computer interaction. It also helps with the activity recognition. As the development of this process, it can apply to a lot of industries.

COVID-19, a serious disease that had a negative effect globally. It still affects our life now. Business has been shut down, School changes to online mode, people work at home. However, this situation does bring the world to notice how important online corporations are to us. Especially for school and universities, other contents can simply upload it online for teaching and marking. But not all classes are suitable to move online, such as PE class. The development of human pose estimation (recognition) is a great tool to help online PE teaching.

2. RELATED WORK

Human pose recognition has a wide application space in computer vision, but at the same time, it is very challenging research work. The significant development of 2D human pose recognition started from the CPM researched by the Yaser's group of CMU in 2016, to the Hourglass proposed by the Deng's group in Princeton, then is the OpenPose that the CMU team present in the COCO, followed by the CPN (2017) and MSPN (2018) that Megvii Technology proposed in the COCO, and the latest HRNet in 2019 of the MSRA Wang's group(Yu, 2019). With the development of human pose recognition technology, it can be used not only for motion detection in teaching, games, or sports events, but also for life state detection, and it can even be used to train robots to make human-like actions.

3. SIGNIFICANCE

3.1 Market needs

During COVID-19, the importance of online corporations has come to the attention of the whole world. Traditional industry has to innovate and adapt the way of online access and development of both end, real life and online. This application would help all schools to teach PE class online. The traditional way is they just post videos online and assume all students would follow the video. This application enables students to finish their daily exercise task with a computer PE teacher. The application can tell if their gesture is wrong or not and tell them how to use a standard gesture to maximum results. It also enables them to upload video and get a mark of each gesture.

3.2 User experience

In our application, users will feel like they have a private teacher to teach them. All they need is the internet and a camera. Few years ago, the same function application needed more equipment. For example, 360 kinect. Without the kinect, the computer can not detect your real life poseture. With the new application, users can use it wherever they are. Simply open the application and start learning.

3.3 Technology innovation

The human pose recognition is a new algorithm to detect human movement and capture human gestures. So computers can easily distinguish human movement from various backgrounds. CMU published a project named “open pose”. This will help our application to recognize human movement. In addition, we can analyze the pose to see if users are doing the right thing.

3.4 The Relationship Between Our Application and Others

Currently, there are no applications that do the same thing like us. Similar application teaching users to do exercise like KEEP. It is an application teaching user exercise pose and class on a mobile app. However, they do it in a traditional way - posting videos and counting time for users. You never find out if you did it right or wrong. Sometimes you may end up achieving nothing because of your wrong pose. Our application is more than that. It can capture your pose and tell you if it is wrong or right, and how you can improve it. This can a lot improve your efficiency on online exercise.

4. REQUIREMENTS

4.1 Hardware and Software

Human pose recognition applications require some basic hardware and specific software. In terms of hardware, a computer that can be connected to the network and a camera are used to collect videos of users' movements, and the required software is composed of human pose recognition software and comparison and score software.

4.2 Platform

By the limitation of given time and cost, the application will be built on the Web. In addition, the Web open-source framework provided on the Internet makes it easier and more flexible to manage functions during application development. Therefore, this application will be developed using Python and the QUI framework.

4.3 User Interface

QUI uses the latest browser technology and provides stylish typographic styles, which can be used as a reference for the interface design and layout of this system. This makes the application more user-friendly and easy to understand.

4.4 Feature

The purpose of this project is to build an application that can use a computer to realize human pose recognition and implement some subsequent functions. The main phases of the system include:

• The first step is to detect people in the video. At this stage, the application detects the key points of the human body in the camera to determine the gestures made by the user. Of course, at the same time, the user can choose to record or not.

• In the next step, the application will use related algorithms and technologies to evaluate whether the user's actions meet the set standards by comparing the collected human poses with the set standards.

• After the action recording/evaluation, the system will generate a score based on the evaluation from this process. Users can see their marks and recording (if there is a recording) in real-time to determine whether to submit for saving or re-recording.

5. SOLUTIONS

The design is to identify the video stream of the user and get the action dataset, then compare it with the standard action data to judge the difference and give a score. In terms of the solutions, the application addresses OpenPose which is based on convolutional neural networks (CNN) and supervised learning to realise realtime pose estimation.

Furthermore, through capturing the key points of the human body, the included angle of the movement is calculated and compared with the standard angle of the movement, so as to achieve the judgment and scoring of the human body posture.

5.1 Human pose estimation

The technology is to estimate human posture by correctly connecting the key points of the human body detected in an image or video. There are some excellent applications of human pose estimation, which are widely used in the fields such as motion recognition, animation and games.

The key points of the human body usually correspond to the joints with a certain degree of freedom on the human body, such as the neck, shoulder, elbow, wrist, waist, knee and ankle (Lin and Lee, 2020). Then, by calculating the relative position of the key points of the human body in the three-dimensional space, the current posture of the human body is estimated, as shown in the Figure 5-1 (Babu, S. C., 2019)below.

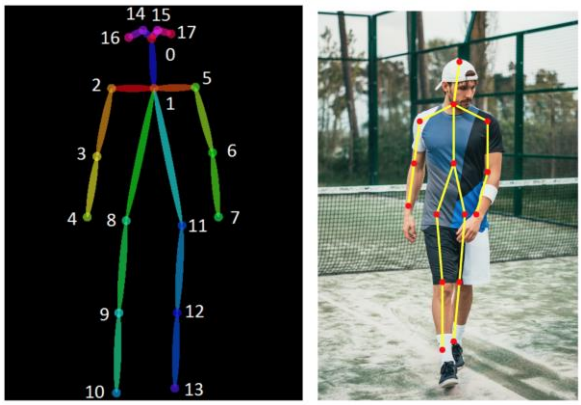

Figure 5-1. Key points of pose estimation sample

However, there are several challenges for attitude detection: Firstly, some external factors such as environmental occlusion, clothing, and changes in lighting make detection difficult. Secondly, the number of people in each picture is unknown and the interaction between people is very complex, such as contact and occlusion, which makes it difficult to combine the various limbs, that is, to determine the parts of a person. Accordingly, the computational amount is positively correlated with the number of people. That is to say, the more people in the image, the greater the computational complexity, which makes real-time detection difficult.

There are different approaches to deal with the situation. Among them, the classical approach uses the pictorial structures framework, but it achieves at the cost of limited expressiveness and does not consider the global context. By contrast, we choose the solution as deep learning based approaches.

5.2 OpenPose

OpenPose is an open source library based on CNN and supervised learning and written with caffe as the framework. It can track people's facial expressions, torso, limbs and even fingers. It is suitable for single or more people with excellent robustness. OpenPose is described as the first real-time 2D attitude estimation of multi-person based on deep learning in the world, which is a milestone in human-computer interaction and provides a high-quality information dimension for the machine to understand humans. The OpenPose's mechanism is shown in the following Figure 5-2 (Zhe Cao, Tomas Simon, Shih-En Wei, & Yaser Sheikh, 2017).

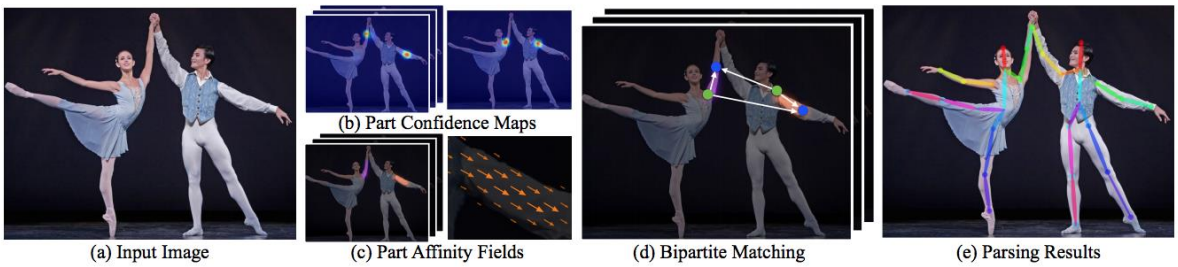

Figure 5-2. Overall pipeline of OpenPose

From above, the process of openpose is mainly divided into three steps:

Firstly, input an image, extract features through the convolutional network, and get a set of feature images; then, divide it into two branches, and extract Part Confidence Maps and Part Affinity Fields (PAF) respectively using CNN networks.

After obtaining these two information, Bipartite Matching is used for finding the Part Association and connecting the relevant nodes of the same person. Due to the vectorial property of PAF itself, the generated Bipartite Matching is quite accurate and is finally merged into the whole skeleton of a person.

The final step is to solve the multi-person Parsing based on PAFs, which can turn the problem into a graph problem through the Hungarian algorithm.

For the detection of human posture, OpenPose uses the bottom-up idea, that is, to find out all the hands and feet in the figure, and then use the matching method to assemble a human skeleton. One drawback to this approach is that there is no way to take advantage of the global contextual information (Yang, Y., & Ramanan, D., 2013).

As a result, PAF overcomes this problem by encoding 2D vectors of limb position and direction in the image field. At the same time, the Part Detection Confidence Map marks the confidence of each key point. Through the two branches, the model can learn the location of the key points and the relationship between them. The model structure is shown below as Figure 5-3 (Zhe Cao, et all., 2017).

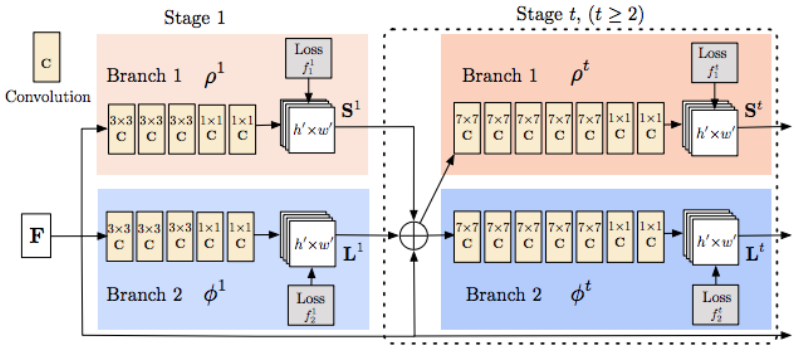

Figure 5-3. Structure of two-branch multi-stage CNN

By extrapolating these bottom-up detection and correlation approaches, the greedy parsing algorithm is used to encode enough of the global context to produce high-quality results at a fraction of the computational cost. In the parallel case, real-time is basically achieved. Moreover, the time consumption is not strongly related to the number of people in the picture.

For a remote sports teaching platform, OpenPose is undoubtedly very suitable for capturing the user's motion. In daily life, just based on 2D images and cheap ordinary cameras, real-time and robust extraction of human bone joints and posture detection can be realized. Photos, videos and webcam video streams that are commonly used in daily life can be used as input, which greatly facilitates the user's operation, which is also an important consideration for us to apply this technology.

5.3 Action rating system

Based on the pose estimation captured by OpenPose, we are working on a set of algorithms that can measure standards of the action. The idea is to calculate the angle of left and right arms, left and right shoulders, left and right legs and so on through the key points of the human body, and then compare it with the standard angle and the allowable error range. Then we can achieve the judgment and scoring of the pose . Besides the judging and rating functions, if possible, we are expecting to provide the rendering and action analysis diagram based on the original image, convenient for the user to analyze.

As for the implementation of the algorithm, we use cosine to calculate the included angles (TensorFlow, 2018), and then calculate the ratio of the number of angles in accordance with the standard to the total number of recognized angles to achieve the evaluation score, as Figure 5-4 below.

$$Evaluation\ Score\ = \frac{The\ number\ of\ qualified\ angles}{Total\ number\ of\ recognized\ angles} \times 100$$

Figure 5-4. Formula of the evaluation score

For the rendering and analysis diagrams, the key points with different confidence degrees are displayed in different colors on the diagrams, those with a score above 0.9 are green, those with a score between 0.5 and 0.9 are yellow, and those below 0.5 are red, which is convenient for user identification. In addition, the original image was rendered to show the key points and main lines, including the body, limbs, head, and face with an additional central axis from the top of the head to the nose and neck. As the following Figure 5-5 shows:

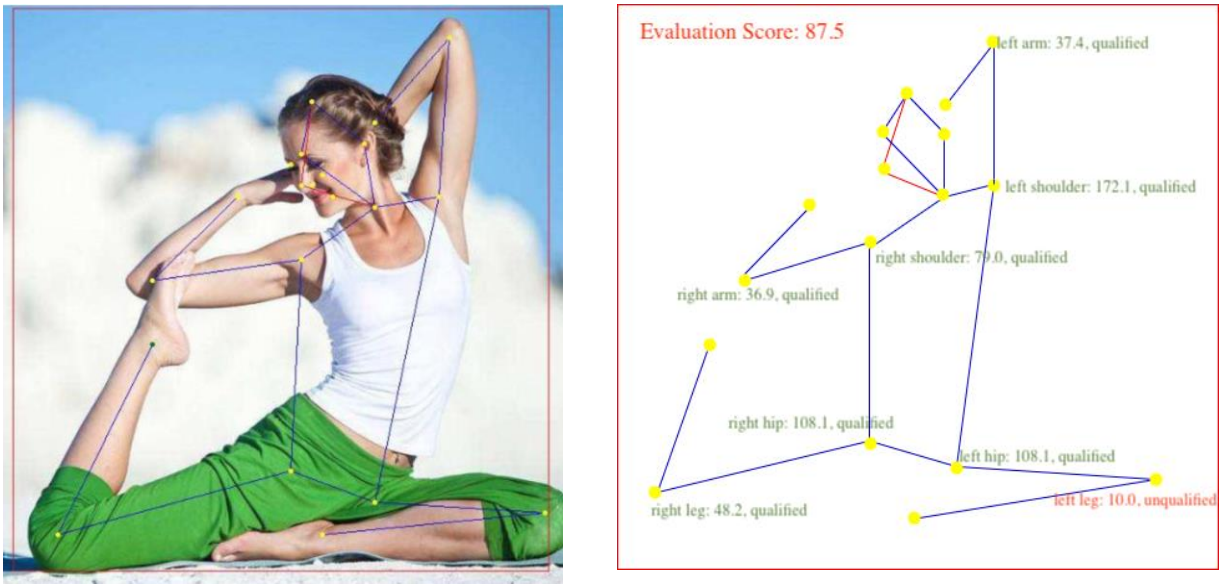

Figure 5-5. Sample rendering and analysis diagram

The new generated image shows key points and major lines, including the central axis running from the top of the head to the nose to the neck, allowing users to see the direction of the face. At the same time, it shows whether the included angle of each position meets the requirements. If the angle is qualified, the description shows green, if not, it shows red. These allow the user to easily analyze the state of the action.

6. PLAN

6.1 Roles & Responsibilities

There are three members in our group and the individual roles and responsibilities are shown in the table below:

Names	Roles	Responsibilities
Yiwen Gong	Leading developer	- Defining requirements - Solution design and function implement - Algorithms and application optimizations
Qianning Gong	Analyst and tester	- Test design and execution - Problem analysis - Data management
Ruozhu Lin	Administer and presenter	- Implement management plans - Manage relevant documents - Organize the presentation

6.2 Milestones & Schedule

The three-layers work breakdown structure consists of milestones and tasks as shown in the table below.

Milestone	Tasks	Deliverables	Date
Week-5	1. Clarifying and assigning individual tasks. 2. Identify the project requirement as well as the implementation method. 3. Building overall blocks.	Overall plan, relevant supporting literature，documents and project proposal draft.	24/03/2020
Week-6	Proposal Report Due	Proposal Report	31/03/2020
Week-7	1. Developing pose estimation function by implementing OpenPose. 2. Debugging parameters.	Realising basic pose estimation function.	07/04/2020
Week-8	1. Implementing a scoring algorithm. 2. Developing the function of generating rendering and analysis diagrams.	Realising scoring, rendering and analysis diagrams.	21/04/2020
Week-9	1. Connecting each function module. 2. Developing extended applications.	Coordination of the overall functions, the basic completing of the application.	28/04/2020
Week-10	1. Deployment 2. Testing	Basically completed applications, video data for testing.	05/05/2020
Week-11	1. Collating documents. 2. Optimizing the application. 3. Preparing final presentation.	Completed application and final related documents, slides for presentation.	12/05/2020
Week-12	Final Presentation and Final Project Report	Relevant Application and Documents	19/05/2020

7. REFLECTION ON WRITING

Babu, S. C. (2019, August 5). A 2019 guide to Human Pose Estimation with Deep Learning. Retrieved from https://nanonets.com/blog/human-pose-estimation-2d-guide/

Cao, Zhe, Simon, Tomas, Wei, & Yaser. (2017, April 14). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Retrieved from https://arxiv.org/abs/1611.08050

Kantaros, D., et al. (2020), Managing the Commercial Impact of the Coronavirus: Implications for the Technology Industry, Retrieved from https://www.natlawreview.com/article/managing-commercial-impact-coronavirus-implications-technology-industry

Lin, J. and Lee, G., 2020. Trajectory Space Factorization For Deep Video-Based 3D Human Pose Estimation. [online] arXiv.org. Available at: <https://arxiv.org/abs/1908.08289>.

Shotton, J., et al.(2011), Real-time human pose recognition in parts from single depth images, *CVPR* , pp. 1297-1304.

Sigal, L. Human pose estimation. Retrieved from https://www.cs.ubc.ca/~lsigal/Publications/SigalEncyclopediaCVdraft.pdf

TensorFlow. (2018, September 27). Move Mirror: An AI Experiment with Pose Estimation in the Browser using TensorFlow.js. Retrieved from https://medium.com/tensorflow/move-mirror-an-ai-experiment-with-pose-estimation-in-the-browser-using-tensorflow-js-2f7b769f9b23

Yang, Y., & Ramanan, D. (2013). Articulated Human Detection with Flexible Mixtures of Parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(12), 2878–2890. doi: 10.1109/tpami.2012.261

Yu, G. (2019). Past, present, future of human pose estimation. Retrieved 30 March 2020, from https://zhuanlan.zhihu.com/p/85506259