



## Data Article

## Tennis player actions dataset for human pose estimation



Chun-Yi Wang<sup>a</sup>, Kalin Guanlun Lai<sup>b</sup>, Hsu-Chun Huang<sup>c</sup>,  
Wei-Ting Lin<sup>d,\*</sup>

<sup>a</sup> Office of Physical Education, National Taichung University of Science and Technology, Taichung, Taiwan

<sup>b</sup> Department of Computer Science & Information Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

<sup>c</sup> Department of Mechanical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

<sup>d</sup> Physical Education Office, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

## ARTICLE INFO

## Article history:

Received 19 May 2024

Revised 10 June 2024

Accepted 17 June 2024

Available online 22 June 2024

Dataset link: [Tennis Player Actions Dataset for Human Pose Estimation \(Original data\)](#)

## Keywords:

Human posture recognition

Pose estimation

Keypoint detection

Tennis action

COCO

Sports Technology

## ABSTRACT

Tennis is a popular sport, and integrating modern technological advancements can greatly enhance player training. Human pose estimation has seen substantial developments recently, driven by progress in deep learning. The dataset described in this paper was compiled from videos of researchers' friend playing tennis. These videos were retrieved frame by frame to categorize various tennis movements, and human skeleton joints were annotated using COCO-Annotator to generate labelled JSON files. By combining these JSON files with the classified image set, we constructed the dataset for this paper. This dataset enables the training and validation of four tennis postures, forehand shot, backhand shot, ready position, and serves, using deep learning models (such as OpenPose). The researchers believe that this dataset will be a valuable asset to the tennis community and human pose estimation field, fostering innovation and excellence in the sport.

© 2024 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

\* Corresponding author.

E-mail address: [titan08148@nkust.edu.tw](mailto:titan08148@nkust.edu.tw) (W.-T. Lin).

Specifications Table

Subject	Computer Science / Computer Vision and Pattern Recognition;
Specific subject area	Data Science / Applied Machine Learning
	Human Posture Recognition;
	Action Recognition;
	Pose Estimation;
	Keypoint Detection
Data format	Filtered
Type of data	.jpeg file (the images from video's frame)
	.json file (COCO-format)
Data collection	The dataset comprises 4 different actions in tennis, each action has 500 images and a COCO-format JSON files.The actions in this dataset, and the action categories name in COCO-format is in brackets:  <ol style="list-style-type: none"><li>1. backhand shot (backhand)</li><li>2. forehand shot (forehand)</li><li>3. ready position (ready_position)</li><li>4. serve (serve)</li></ol> We organize two main directories: annotations and images.  <ul style="list-style-type: none"><li>• annotations: the JSON files of the actions (COCO-format)<ul style="list-style-type: none"><li>■ We use COCO-Annotator to annotating and categorizing human actions.</li></ul></li><li>• images: the images of the actions (according four actions classify to four folders)<ul style="list-style-type: none"><li>■ The images in the dataset were extracted frame by frame from videos that were self-recorded, and manually classified according to different tennis actions.</li></ul></li></ul>
Data source location	Taipei Tennis Center, in Taipei City, Taiwan.
Data accessibility	Repository name: Tennis Player Action Dataset for Human Pose Estimation Data identification number: <a href="https://data.mendeley.com/datasets/nv3rpsxhkhk">10.17632/nv3rpsxhkhk</a> Direct URL to data: <a href="https://data.mendeley.com/datasets/nv3rpsxhkhk">https://data.mendeley.com/datasets/nv3rpsxhkhk</a>

1. Value of the Data

- This dataset has significantly contributed to sports technology by integrating computer vision techniques to further the advancement of sports tech.
- Employing the widely used COCO-format and annotates human skeletal joints (key points), facilitating easy access and training for users.
- The dataset is meticulously curated to capture the nuances of tennis movements, providing detailed annotations for a variety of actions such as serves, volleys, and groundstrokes. This allows for the development of precise pose estimation models that are highly effective in analyzing and enhancing tennis performance.
- If needed, users can also utilize this dataset for other applications, such as tracking tennis balls, by labeling and training it on their own.

2. Background

Datasets related to sports provide valuable data for a wide range of research fields, including policy-making, education, public health, and sports science. Traditionally, these datasets mainly contain raw statistics on athletes' physical conditions, outputs from various modeling efforts, or data collected through software tools, all of which contribute significantly to the advancement of these fields [1-4]. As technology progresses, computer vision has emerged as a critical area of research, especially in human pose estimation. This domain has witnessed the development of specific datasets to support such research efforts [5-6]. Consequently, several datasets

within the field of sports science are now specifically designed for training human pose estimation models. For instance, LDCNet focuses on flexible human pose estimation by leveraging limb direction cues, highlighting its application in industrial behavioral biometrics systems [7]. Another example is ARHPE, which employs asymmetric relation-aware representation learning to enhance head pose estimation, a crucial aspect in industrial human-computer interaction [8]. Additionally, MFDNet advances the field by integrating collaborative pose perception with matrix Fisher distribution for precise head pose estimation [9]. These datasets are integral in supporting the development and refinement of human pose estimation models, facilitating advancements in sports science and other related fields.

Traditional practices for creating datasets have predominantly relied on computer vision techniques. However, with the advent and evolution of deep learning, the need for raw image files and annotation data has become paramount. The COCO format has emerged as a standard for annotation data in recent years [8]. Our dataset also utilizes the COCO format, facilitating its use for training and validation in deep learning models such as OpenPose [9] and MediaPipe [10]. This dataset comprises a comprehensive collection of annotated tennis action images, designed to train models capable of recognizing specific postures within tennis matches. Moreover, users can customize the dataset for other applications, such as tracking tennis balls, by adding their own labels and training data. The primary objective of this dataset is to promote the advancement of computer vision applications in tennis-related fields. In recent years, several significant advancements have been made in the field of computer vision and deep learning. For example, OpenPose represents a notable improvement over traditional pose estimation models by providing a multi-person pose detection framework, which enhances the accuracy and applicability of pose recognition [9]. Similarly, MediaPipe has extended these capabilities by integrating real-time pose estimation with high efficiency, making it a preferred tool for a wide range of applications beyond sports [10]. The progression from basic pose estimation models to more sophisticated frameworks illustrates the rapid development and increasing precision of computer vision technologies, which our dataset aims to support.

### 3. Data Description

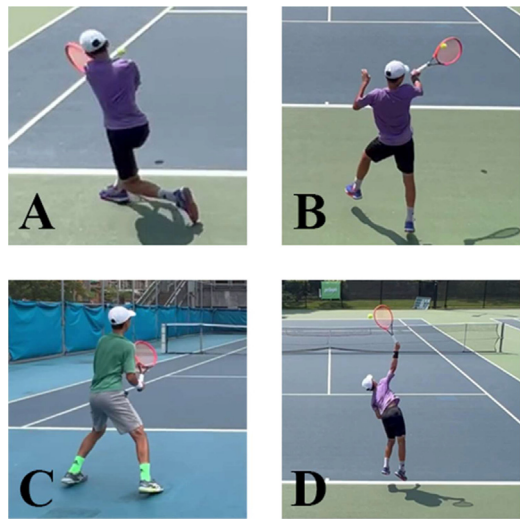
This dataset is designed for human pose estimation applications within tennis, featuring commonly observed tennis postures including forehand stroke, backhand stroke, ready position, and serve, as shown in Fig. 1.

This dataset contains two parts: 1. images from the frame of the video of the players' action, and 2. the action annotation JSON files (COCO-format). Part 1 have 2,000 images, part 2 have 4 files, and it on Mendeley Data shown in Fig. 2 (size on disk is about 508 MB (533,372,928 bytes)).

The researchers organized two parts as two main directories. One is images, it divided into four subfolders by posture in "images" folder. The files in the subfolders are named by researchers, following a specific convention. Researchers extract the first letter of the parent folders and assign sequential numbering. For instance, the name of the images within "images/backhand" folder have prefix "B\_", and followed by a numerical sequence (e.g., B\_001, B\_002, ..., B\_500). The Other is annotation JSON files, it has four files and named by four postures. Folder structure is shown in Fig. 3.

### 4. Image Information

The dataset described in this article contains 500 images for each posture, total have 2000 images. Before classified to four specific actions, the researchers first recorded videos of themselves playing tennis, then analyzed these videos' frame by frame to classify the frames' image into the specified actions. These videos were captured using a smartphone with a resolution of



**Fig. 1.** Common tennis postures. (A) Backhand stroke, (B) Forehand stroke, (C) Ready Position, (D) Serve.

**Table 1**  
Brief description about the data collection.

No.	Particulars	Description
1	Data type	4 tennis postures:  (1) forehand stroke (2) backhand stroke (3) ready position (4) serve
2	Original data format	Video file using H.264/MPEG-4 AVC codec (.mp4) Resolution: 720P (1280 × 720 pixels) Frame Rate: 30 fps
3	Filtered data format	JPEG image file (.jpeg) Resolution: 720P (1280 × 720 pixels)
4	Period and Date	January-December 2023
5	Participants	Member of the World Junior Team Championships (Lin,yu-min. Taiwan)
6	Location	Taipei Tennis Center

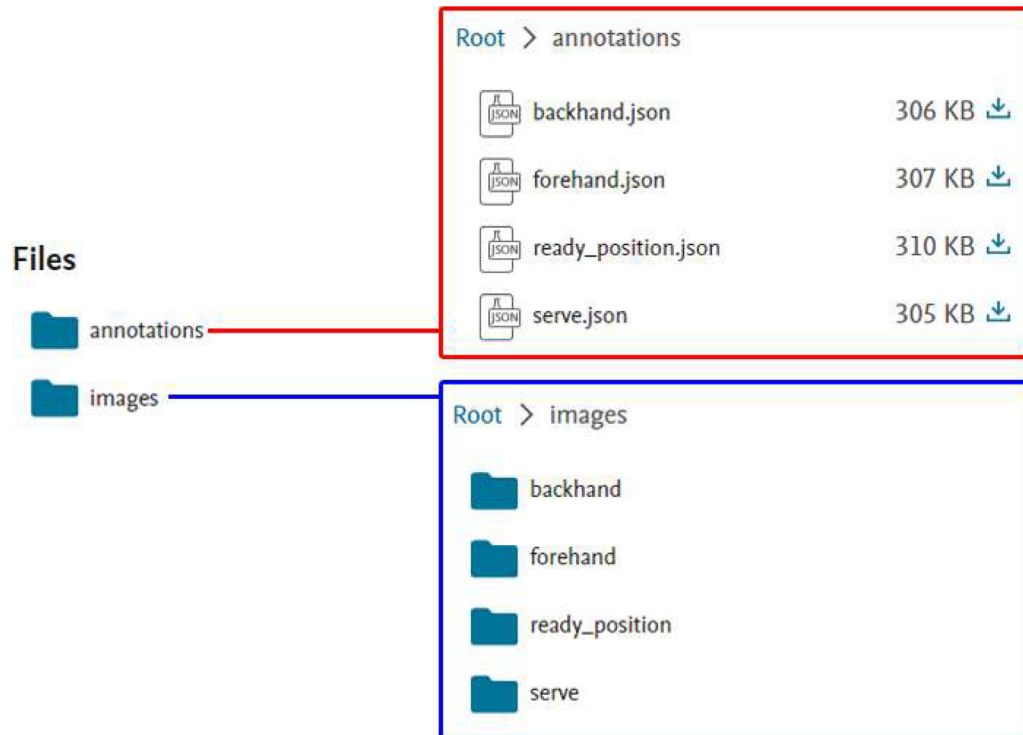
Source: Author's own organization.

720P, with dimensions of 1280 pixels width and 720 pixels height, and a frame rate of 30 fps, so the images' resolution also is 1280 × 720. The data collection outline is shown in Table 1.

5. Annotation File Information

The researchers utilize the extracted images to annotate human skeletal joints and classify the postures. For supplying common deep learning model of human pose estimation to train, the researchers use COCO-Annotator [11] as annotation tool, annotated joints are illustrated in Fig. 4 and the joints number and name pairs are shown in Table 2.

Because the researchers use COCO-Annotator, so the generated JSON files must be COCO format. The JSON file's format in this dataset is shown in Table 3. The instances of categories' name, keypoints, and skeleton in JSON file is represent to Fig. 5, the name of category in the red box is the posture's name, the category information is set by researchers using COCO-Annotator. The instance of image and annotation in the JSON file is represent to Fig. 6.



**Fig. 2.** Dataset files on Mendeley Data.

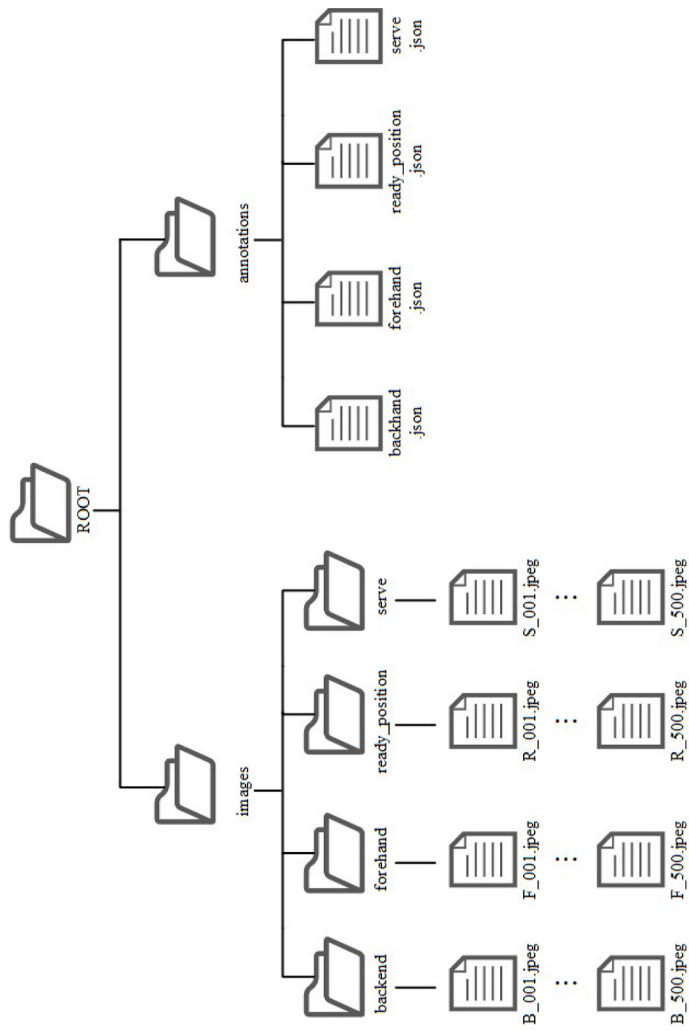


Fig. 3. Folder structure.

**Table 2**

Skeleton joint number to name pairs in Fig. 4.

Joint's No.	Joint's Name
0	nose
1	left_eye
2	right_eye
3	left_ear
4	right_ear
5	left_shoulder
6	right_shoulder
7	left_elbow
8	right_elbow
9	left_wrist
10	right_wrist
11	left_hip
12	right_hip
13	left_knee
14	right_knee
15	left_ankle
16	right_ankle
17	neck

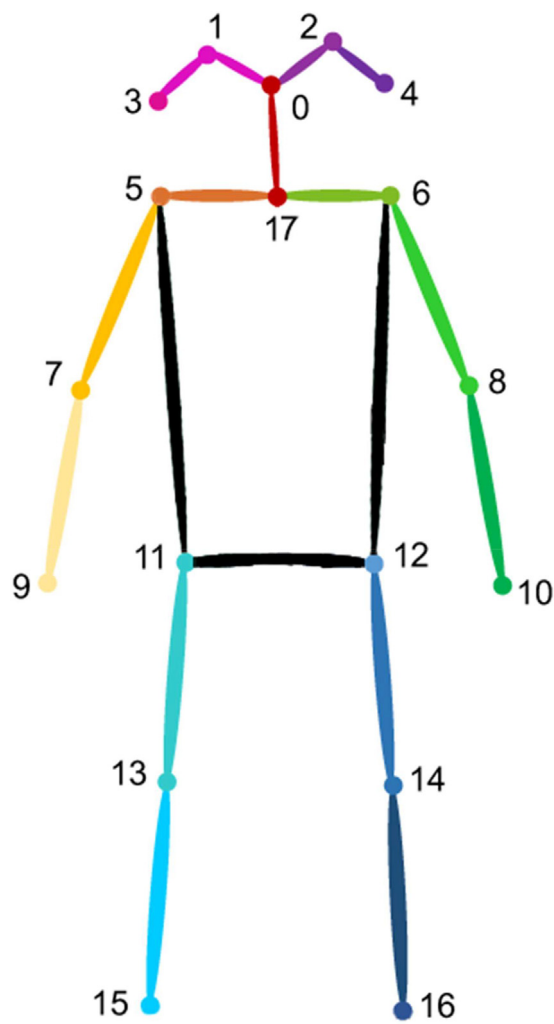
Source: Author's own organization.

**Table 3**

JSON file's format of this dataset.

<pre> {   "images": [     {       "id": (image ID, same as the id in part of "annotations"),       "dataset_id": (dataset ID),       "path": (image file path),       "width": (image width),       "height": (image height),       "file_name": (image file name)     },     ...   ],   "categories": [     {       "id": (category ID),       "name": (category name, same as the id in part of "annotations"),       "keypoints": (keypoints list),       "skeleton": (skeleton (connected keypoints) list)     }   ],   "annotations": [     {       "id": (annotation ID),       "image_id": (image ID, same as the id in part of "images"),       "category_id": (category ID, same as the id in part of "category"),       "segmentation": (polygon list),       "area": (the area of the target box. unit: pixel),       "bbox": (the coordinates list of the target box' each corners),       "iscrowd": (whether the image is a crowd),       "isbbox": (whether the image has target box),       "keypoints": (list of keypoints coordinates on the image),       "num_keypoints": (number of keypoints on the image)     },     ...   ] } </pre>
--

Source: Author's own organization.



**Fig. 4.** Human skeleton joints in the dataset.

**6. Experimental Design, Materials and Methods**

The camera is positioned at the rear of the tennis court, capturing the player from behind (aligned with the player's facing direction). The camera is positioned approximately 6.4 meters from the court's baseline. The height, while not fixed or recorded, is considered by researchers to be inconsequential for the purpose of analyzing tennis movements. The location of the camera setup is shown in Fig. 7.

Before playing tennis, we set up an experimental recording environment like above. Once the setup is complete, we began recording the video. Subsequently, we retrieve the video frame by frame, classifying each frame according to the specific tennis action it captures to compile an image dataset. Then the researchers annotated target human's skeleton joint for each image in these images using COCO-Annotator and generate labeled JSON files to construct the dataset for this article. The process flow is illustrated in Fig. 8.

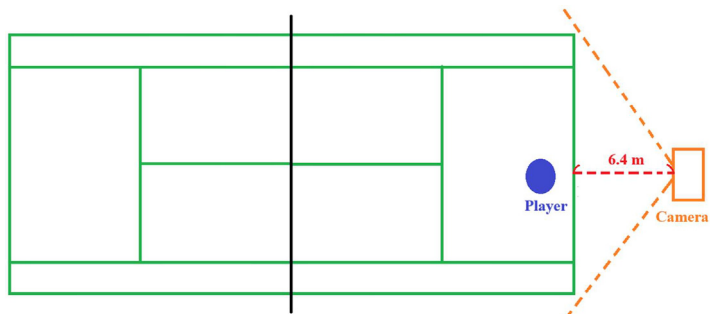


```
categories: [
{
  "id": 5,
  "name": "Backhand",
  "keypoints": [
    "nose",
    "left_eye",
    "right_eye",
    "left_ear",
    "right_ear",
    "left_shoulder",
    "right_shoulder",
    "left_elbow",
    "right_elbow",
    "left_wrist",
    "right_wrist",
    "left_hip",
    "right_hip",
    "left_knee",
    "right_knee",
    "left_ankle",
    "right_ankle",
    "neck"
  ],
  "skeleton": [
    [1, 2],
    [1, 3],
    [1, 10],
    [2, 4],
    [3, 5],
    [6, 8],
    [6, 12],
    [6, 18],
    [7, 9],
    [7, 13],
    [7, 18],
    [8, 10],
    [9, 11],
    [12, 14],
    [12, 13],
    [13, 15],
    [14, 16],
    [15, 17]
  ]
},
],
categories: [
{
  "id": 6,
  "name": "Forehand",
  "keypoints": [
    "nose",
    "left_eye",
    "right_eye",
    "left_ear",
    "right_ear",
    "left_shoulder",
    "right_shoulder",
    "left_elbow",
    "right_elbow",
    "left_wrist",
    "right_wrist",
    "left_hip",
    "right_hip",
    "left_knee",
    "right_knee",
    "left_ankle",
    "right_ankle",
    "neck"
  ],
  "skeleton": [
    [1, 2],
    [1, 3],
    [1, 10],
    [2, 4],
    [3, 5],
    [6, 8],
    [6, 12],
    [6, 18],
    [7, 9],
    [7, 13],
    [7, 18],
    [8, 10],
    [9, 11],
    [12, 14],
    [12, 13],
    [13, 15],
    [14, 16],
    [15, 17]
  ]
},
],
categories: [
{
  "id": 7,
  "name": "Ready Position",
  "keypoints": [
    "nose",
    "left_eye",
    "right_eye",
    "left_ear",
    "right_ear",
    "left_shoulder",
    "right_shoulder",
    "left_elbow",
    "right_elbow",
    "left_wrist",
    "right_wrist",
    "left_hip",
    "right_hip",
    "left_knee",
    "right_knee",
    "left_ankle",
    "right_ankle",
    "neck"
  ],
  "skeleton": [
    [1, 2],
    [1, 3],
    [1, 10],
    [2, 4],
    [3, 5],
    [6, 8],
    [6, 12],
    [6, 18],
    [7, 9],
    [7, 13],
    [7, 18],
    [8, 10],
    [9, 11],
    [12, 14],
    [12, 13],
    [13, 15],
    [14, 16],
    [15, 17]
  ]
},
],
categories: [
{
  "id": 8,
  "name": "Serve",
  "keypoints": [
    "nose",
    "left_eye",
    "right_eye",
    "left_ear",
    "right_ear",
    "left_shoulder",
    "right_shoulder",
    "left_elbow",
    "right_elbow",
    "left_wrist",
    "right_wrist",
    "left_hip",
    "right_hip",
    "left_knee",
    "right_knee",
    "left_ankle",
    "right_ankle",
    "neck"
  ],
  "skeleton": [
    [1, 2],
    [1, 3],
    [1, 10],
    [2, 4],
    [3, 5],
    [6, 8],
    [6, 12],
    [6, 18],
    [7, 9],
    [7, 13],
    [7, 18],
    [8, 10],
    [9, 11],
    [12, 14],
    [12, 13],
    [13, 15],
    [14, 16],
    [15, 17]
  ]
},
],
```

**Fig. 5.** The instance of the categories' part in JSON file (for 4 postures).

[illegible]

**Fig. 6.** The instance of the part of image and annotation in JSON file (from 2 backhand images).



**Fig. 7.** Location of camera setup in the experimental field.

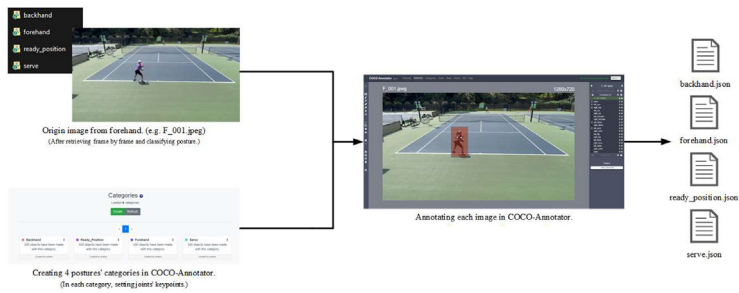


Fig. 8. The process flow of annotating using COCO-Annotator.

Limitations

When creating this dataset, the camera setup was not at a fixed distance; it was simply positioned behind the player to record the various strokes. The dataset currently includes only four actions: forehand shot, backhand shot, ready position, and serve. While these encompass most of the essential tennis movements, some minor actions might still be missing. In the future, we aim to expand the collection to include a broader range of movements and to augment the dataset with additional images.

Ethics Statement

After the performance of the experiment, we blurred all unrelated people in images (the people of the opposite field). And the participant (the person back on the camera) in images is the authors' friend, he provided some data related to physical status and habits of individual, and they read and signed an informed consent form, conserved at Physical Education Office at "National Kaohsiung University of Science and Technology" (the correspondent's office). We follow Research ethics guidelines in everything we do, and have obtained a certificate from the local Center for Taiwan Academic Research Ethics Education, certificate number: P107259575-1.

Data Availability

[Tennis Player Actions Dataset for Human Pose Estimation \(Original data\)](#) (Mendeley Data).

CRedit Author Statement

**Chun-Yi Wang:** Conceptualization, Data curation, Methodology; **Kalin Guanlun Lai:** Software, Writing – original draft; **Hsu-Chun Huang:** Validation, Writing – review & editing; **Wei-Ting Lin:** Supervision, Validation, Writing – review & editing.

Acknowledgement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] D.I. Bourdas, P. Bakirtzoglou, A.K. Travlos, V. Andrianopoulos, E. Zacharakis, Analysis of a comprehensive dataset: Influence of vaccination profile, types, and severe acute respiratory syndrome coronavirus 2 re-infections on changes in sports-related physical activity one month after infection, *Data Brief* 51 (2023) 109723.
- [2] C. Mountfield, Data on Gaussian copula modelling of the views of sport club members relating to community sport, Australian sport policy and advocacy, *Data Brief* 42 (2022) 108111.
- [3] P. Pinheiro, L. Cavique, Regular sports services: dataset of demographic, frequency and service level agreement, *Data Brief* 36 (2021) 107054.
- [4] W. Limroongreungrat, C. Mawhinney, S. Kongthongsung, C. Pitaksathienkul, Landing error scoring system: data from youth volleyball players, *Data Brief* 41 (2022) 107916.
- [5] N.R.P. Rodrigues, N.M.C.da Costa, R. Novais, J. Fonseca, P. Cardoso, J. Borges, AI based monitoring violent action detection data for in-vehicle scenarios, *Data Brief* 45 (2022) 108564.
- [6] A.V. Ruescas-Nicolau, E.J. Medina-Ripoll, E.P. Bernabé, H. Martínez, R. de, Multimodal human motion dataset of 3D anatomical landmarks and pose keypoints, *Data Brief* 53 (2024) 110157.
- [7] Y. Suryawanshi, N. Gunjal, B. Kanorewala, K. Patil, Yoga dataset: a resource for computer vision-based analysis of Yoga asanas, *Data Brief* 48 (2023) 109257.
- [8] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan D., Zitnick, C. L., and Dollár, P., "Microsoft COCO: common objects in context." (2015) arXiv:1405.0312.
- [9] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, Y. Sheikh, OpenPose: realtime multi-person 2D pose estimation using part affinity fields, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (1) (2021) 172–186, doi:10.1109/TPAMI.2019.2929257.
- [10] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja E., Hays, M., Zhang, F., Chang, C.-L., Yong M., Lee, J., Chang, W.-T., Hua, W., Georg, M., and Grundmann, M., "MediaPipe: a framework for building perception pipelines." (2019) arXiv:1906.08172.
- [11] Brooks, J. "COCO annotator." 2019. Retrieved from <https://github.com/jsbrooks/coco-annotator/>.