Deep Learning Model for Integrated Estimation of Wheelchair and Human Poses Using Camera Images

Shimpei Aihara[1, 2], Takara Sakai[3], Ryusei Shibata[3], Toshiaki Matsubara[3], Ryosuke Mizukami[3], Yudai Yoshida[4] and Akira Shionoya[3]
[1] Department of Sport Sciences, Japan Institute of Sport Sciences, Tokyo, Japan
[2] School of Creative Science and Engineering, Waseda University, Tokyo, Japan
[3] Department of Management and Information Systems Science, Nagaoka University of Technology, Niigata, Japan
[4] Faculty of Liberal Arts, Tohoku Gakuin University, Miyagi, Japan

Introduction: Human pose estimation from camera images is popular with the development of deep learning [1]. In sports, it is a useful tool for motion analysis [2]. It is also possible to be used in wheelchair sports. The wheelchair is important equipment that functions as an athlete's legs. Therefore, wheelchair pose estimation, as well as human pose estimation, is essential. However, the pose estimation of humans with wheelchairs has not been reported in previous literature. This study aimed to develop a deep learning model for integrated estimation of wheelchair and human poses using camera images.

Methods: Firstly, a wheelchair sports dataset was constructed. Royalty-free images of wheelchair sports were collected from the internet and the key point coordinates on the images were manually annotated by sports biomechanics experts. The resolutions of images were VGA (640 × 480 pixels). There were 17 key points defined as follows: 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, center of the left wheel, center of the right wheel, bottom of the left wheel, and bottom of the right wheel. There were approximately 2300 images from about 6000 subjects. The data were randomly divided into training data and test data in a ratio of 7:3. A deep learning model was designed using convolutional neural networks. The architecture of the proposed model is shown in Figure 1. The model was based on MASK R-CNN [3], which is widely used for human pose estimation

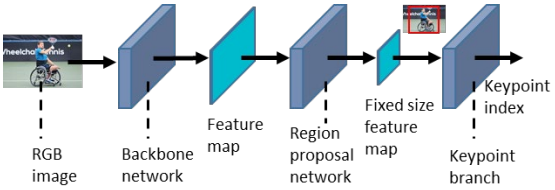

Fig.1 Architecture of the proposed model

Corresponding author email: shimpei.aihara@jpnsport.go.jp

and has no license limitation. Finally, the designed model was trained on a large human pose dataset [4], and then transfer learned the model on a wheelchair sports dataset.

Results: The absolute error between the estimated coordinates and the manually annotated coordinates ware evaluated using test data. For the key points of humans, the absolute error was 4.43 pixels. The widely used methods for human pose estimation, Mask R-CNN [3] and Open Pose [5] 4.68 and 4.51 pixels. The value of the proposed method was greater than that of previous studies [3][5]. The proposed method improved the estimation error by over 1.7% compared with previous studies [3][5]. For the wheelchairs' key points, the absolute error was 6.22 pixels. The results confirmed that the proposed method applied to various types of wheelchair sports, scenes, and individuals, as shown in Figure 2.

Fig.2 Examples of estimation results by the proposed model. These were cropped to focus on wheelchairs and humans.

Conclusion: A deep learning model for integrated estimation of wheelchair and human poses using camera images was developed. The high estimation accuracy was confirmed and it was established that this would be a useful tool for wheelchair sports. Our deep learning approach can be applied for estimating the pose of hand-held or body-worn tools and equipment, such as rackets, bats, and ski boards, in addition to the human pose.

1. Cust EE, Sweeting AJ, Ball K, Robertson S (2019) Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. J Sports Sci 37:568–600.
2. Difini GM, Martins MG, Barbosa JLV (2021) Human pose estimation for training assistance: A systematic literature review. ACM Int Conf Proceeding 189–96.
3. He K, Gkioxari G, Dollar P, Girshick R (2017) Mask R-CNN. Proc IEEE Int Conf Comput Vis 2980–2988.
4. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft COCO: Common objects in context. Eur Conf Comput Vis 8693:740–755.
5. Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y (2021) OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 43:172–186.

Corresponding author email: shimpei.aihara@jpnsport.go.jp