Disclaimer

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

Journal of Biomechanics 123 (2021) 110460

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Using deep neural networks for kinematic analysis: Challenges and opportunities

Neil J. Cronin [*]

Neuromuscular Research Centre, Faculty of Sport and Health Sciences, University of Jyvaskyla, Finland
School of Sport and Exercise, University of Gloucestershire, UK

ARTICLE INFO

Article history:
Accepted 20 March 2021

Keywords:
Motion analysis
Kinematics
Deep neural network
Markerless tracking
AI

ABSTRACT

Kinematic analysis is often performed in a lab using optical cameras combined with reflective markers. With the advent of artificial intelligence techniques such as deep neural networks, it is now possible to perform such analyses without markers, making outdoor applications feasible. In this paper I summarise 2D markerless approaches for estimating joint angles, highlighting their strengths and limitations. In computer science, so-called "pose estimation" algorithms have existed for many years. These methods involve training a neural network to detect features (e.g. anatomical landmarks) using a process called supervised learning, which requires "training" images to be manually annotated. Manual labelling has several limitations, including labeller subjectivity, the requirement for anatomical knowledge, and issues related to training data quality and quantity. Neural networks typically require thousands of training examples before they can make accurate predictions, so training datasets are usually labelled by multiple people, each of whom has their own biases, which ultimately affects neural network performance. A recent approach, called transfer learning, involves modifying a model trained to perform a certain task so that it retains some learned features and is then re-trained to perform a new task. This can drastically reduce the required number of training images. Although development is ongoing, existing markerless systems may already be accurate enough for some applications, e.g. coaching or rehabilitation. Accuracy may be further improved by leveraging novel approaches and incorporating realistic physiological constraints, ultimately resulting in low-cost markerless systems that could be deployed both in and outside of the lab.

1. Introduction

In recent years, the long-held dream of taking biomechanical analyses out of the laboratory has edged closer to reality with the advent of new technology. For example, wearable devices can now be used to track individual stride characteristics during gait (e.g. Davidson et al., 2019; Liu et al., 2010), including joint angles (Mundt et al., 2020; Zimmermann et al., 2018). The computation of joint angles is challenging with wearable devices but can be achieved relatively easily using a set of cameras. Until recently, kinematic analysis was generally performed in a lab using optical cameras in combination with reflective markers, but this setup is not primarily designed for outdoor use (see Colyer et al., 2018 for a review of the methodological development). With the advent of deep neural networks (deep learning; see Table 1 for a glossary of key terms), it is now possible to estimate joint angles without the need for reflective markers. This requires combining one or more cameras with an approach referred to in computer science as "pose estimation" to detect body landmarks, and then using simple geometry to estimate joint angles (i.e. the angle between two vectors that each represent a body segment). Thus, at least in theory, kinematic analysis can be performed outside of a laboratory, including in clinical and sporting environments (Cronin et al., 2019; Kidziński et al., 2020). The purpose of this paper is to summarise popular markerless approaches for estimating joint angles, highlighting their strengths and limitations. I focus mainly on 2D applications, since the use of pose estimation for markerless 3D joint angle prediction is still in its infancy (see Nakano et al., 2020; Nath et al., 2019).

2. Pose estimation

In the past few years, user-friendly neural network-based methods for pose estimation have emerged that allow markerless detection of anatomical landmarks (Arac et al., 2019; Cao et al., 2017;

[*] Address: Viveca 234, Rautpohjankatu 8, 40700 Jyväskylä, Finland.
E-mail address: neil.j.cronin@jyu.fi

https://doi.org/10.1016/j.jbiomech.2021.110460
0021-9290/© 2021 The Author(s). Published by Elsevier Ltd.

Table 1
Glossary of key terms.

Neural network	An iterative computational method that uses a network of functions to learn features from data. Convolutional neural networks are a variant commonly used in image processing because of their use of mathematical convolutions to detect spatial patterns. Graph neural networks are another variant that allow connections between different structures to be encoded in the model, e.g. the spatial relationships between body parts
Deep learning	Using neural networks with multiple layers (hence "deep") to detect patterns in datasets
Ground truth	Known "correct" answers against which a neural network's predictions are compared to determine its accuracy, e.g. joint centre locations in an image determined manually
Overfitting	A process where a trained model learns the features of the training data so well ("overfits" the data) that it does not generalise well (makes poor predictions when exposed to new data)
Pose estimation	A computer vision method that uses some kind of neural network model to detect body landmarks in an image
Probability heat map	Convolutional neural networks assign probabilities to each pixel of an image depending on the learned likelihood that a feature is present in that part of the image. For example, a model trained to detect hands will assign high probabilities to pixels where the hands are visible and low probabilities to other pixels
Self-supervised learning	A method of training a neural network that does not require the user to provide manually labelled data as input. Instead the labels are extracted automatically from the data. For example, we could cut an image into 9 equally-sized squares and jumble them up. By learning to rearrange the squares in the correct order, the model can "learn" useful features from the image
Supervised learning	The process of using labelled data to train a neural network to learn desired features. After training, the network can detect the presence of learned features in new, previously unseen images
Transfer learning	Using a model trained to perform one task as the basis of a model for a new task

Graving et al., 2019; Mathis et al., 2018). Some of these methods even allow videos to be processed in real-time (Cao et al., 2017; Kane et al., 2020). One algorithm that has received particular attention is DeepLabCut (Mathis et al., 2018), which was initially designed for tracking animal behaviour, but can also be used to track human movement in 2D or 3D (Cronin et al., 2019; Nath et al., 2019). These and many other recent studies have demonstrated the potential value of markerless neural network approaches in the field of human movement science (see also Tome et al., 2018). Could these methods lead to a revolution in human motion analysis?

To address this question, it is first important to examine where these new approaches came from. In computer science, the field of pose estimation (Table 1) has existed for many years, and the current state of the art is quite advanced, with several ongoing competitions in this area ensuring continuous development of new methods (e.g. https://posetrack.net/). For example, an open source method called OpenPose enables key body landmarks to be tracked from multiple humans in a video in real-time (Cao et al., 2019, 2017), and has been used as part of a 3D markerless system to calculate joint angles during gait with promising results (Nakano et al., 2020). However, there are some critical distinctions between pose estimation and kinematic analysis. Firstly, strictly speaking pose estimation only involves the detection of body landmarks, which are then used in combination with geometry to compute the angle between any two body segments. Secondly, the accuracy requirements of pose estimation are less strict than those of kinematic analysis. Common applications of pose estimation include gaming, robotics and animation, and these algorithms are also useful to help automated vehicles detect pedestrians. For these applications, it is usually sufficient to predict the location of a body landmark to within about 5–10 cm. However, when calculating joint angles for kinematic analysis, this magnitude of error is unacceptable. Thus, pose estimation algorithms cannot simply be used out of the box for accurate kinematic analysis (see Seethapathi et al., 2019 for further limitations).

Nonetheless, the emergence of new, more advanced approaches that allow a user to train their own models (such as DeepLabCut) may give us the raw ingredients needed to develop markerless deep learning approaches that could contend with existing gold standard methods such as optical motion analysis (and manual digitisation). However, to develop a markerless deep learning method for estimating joint angles, it is first necessary to train a model to detect the desired features, which in this case are anatomical landmarks. Existing pose estimation methods achieve this via a process called supervised learning (see Mathis et al., 2020 for discussion of individual algorithms).

3. Supervised learning

In the context of this paper, supervised learning involves training an algorithm to identify patterns between images and their corresponding labels, which are provided by a human 'supervisor' (Cunningham et al., 2008). These labels indicate where in the image a particular body part or object is located. The premise is that after seeing a sufficient number of examples of a body part's appearance, the network can robustly learn to identify this body part in other images that it has not previously seen (Fig. 1).

Unfortunately, the labelling process is fraught with difficulties. Firstly, it is inevitably subjective. Each labeller has their own concept of anatomical landmarks, and where exactly the label should be placed. If all of the data that are used to train the model (i.e. training data) are labelled by the same person, the network may learn to identify the body parts consistently according to that labeller's logic, but a different labeller may still argue that the neural network labels images incorrectly (Nowak and Rüger, 2010). There is no easy solution to this problem because we often do not know the ground truth (i.e. the objectively correct location), but one approach is for 2 or more people to label the data and then confirm agreement between their estimates based on some predefined reliability criterion.

Another difficulty of the labelling process relates to the quality and variety of the training data. When selecting these data, it is common practice to first collect videos that are relevant to the task at hand, for example, videos of people walking and running. We then extract individual images from those videos and label the extracted images. The goal is to produce a training set that includes lots of variability, so that the neural network learns to label images robustly. In our example, we would want to include images from different parts of the step cycle, people wearing different clothes, with different skin colours, different lighting, and from different angles and scales. By exposing our network to all these sources of variability, there is a better chance that after training it will recognise wide variations in new images. Naturally, these requirements mean that large and varied training sets need to be collated and labelled, both of which can be time consuming. Moreover, if we want a robust markerless approach, we cannot train the model using images that contain reflective markers (which could act as ground truth). If we did, the model could use the appearance of the markers to help identify each body part, so when trying to analyse a new image where the markers were not present, the model may not make accurate predictions.

Camera settings are another issue relevant to the collection of training data, particularly the frame rate and shutter speed with

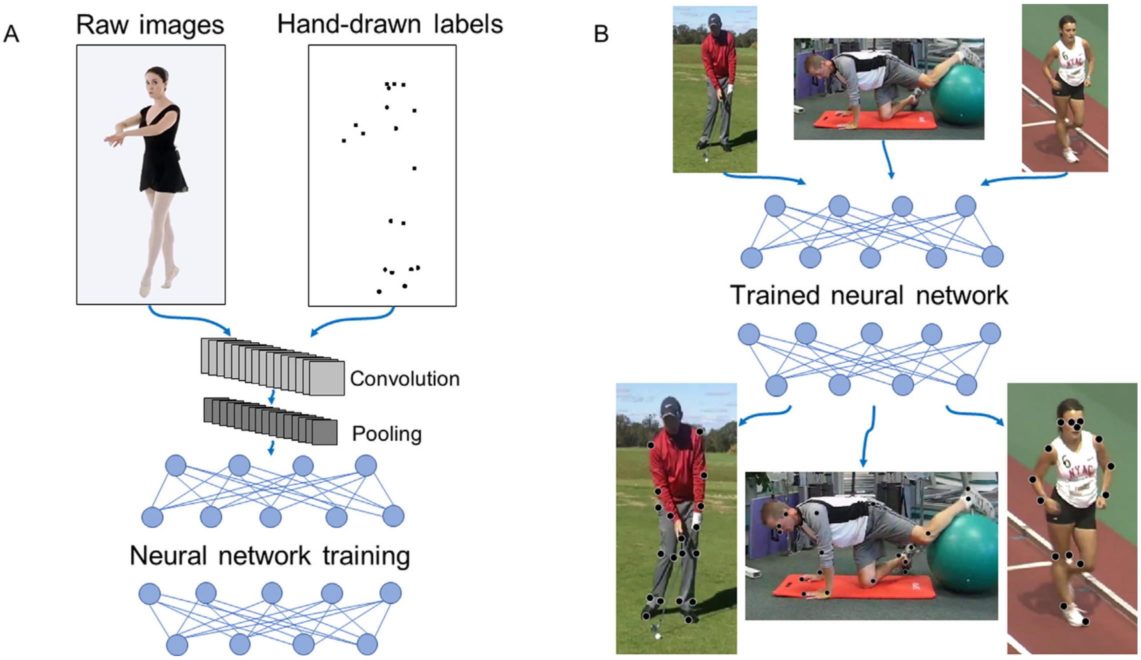

Fig. 1. In supervised learning we first train a convolutional neural network (A, for definition see Table 1) by feeding in image-label pairs (only one pair shown for simplicity). Once the model is trained, we perform inference, i.e. process new images with the trained model. The model labels the images using the logic that was 'learned' during training (B). All images are from the MPII dataset (Andriluka et al., 2014).

which the videos are sampled (Mathis et al., 2020). As a general rule, shutter speed should be at least double the sampling rate. Many modern cameras, such as those in mobile phones and webcams, will by default sample data at around 30 frames per second, necessitating a shutter speed of at least 1/60th of a second. With these settings, the individual frames of a video can be very blurry in dynamic scenes (Fig. 2A), and this makes labelling challenging. Depending on the camera, this can usually be overcome by manually increasing shutter speed. Image resolution is another important factor: very low-resolution images result in pixellated close-up views that can make it difficult to accurately label a body part (Fig. 2B). Furthermore, body part occlusion is common in a 2D camera view, e.g. the hand blocking the hip (Fig. 2C). The labeller must then decide whether to label the location where the blocked part is believed to be, or to avoid labelling the part for that image (see Table 2 for some recommendations).

As well as data quality, data volume is a key element of supervised learning. Generally, neural networks require lots of training examples to reliably identify an object or body part in new images. Even a task as mundane as identifying cats in images is surprisingly difficult for neural networks, requiring tens of thousands of examples (e.g. Krizhevsky et al., 2012). Thus, to train robust models, we need willing individuals to label the training images. In large, open source datasets, this task is achieved using crowdsourcing whereby a large number of individuals are recruited and paid to each label a subset of images (the largest dataset currently in existence, ImageNet, currently contains over 14 million labelled images; Deng et al., 2009). All the labelled data are then combined and used to train one large model that includes thousands of labelled images. A good example is OpenPose (Cao et al., 2019, 2017), which was trained using tens of thousands of images that were labelled by a large number of people. This is problematic because we cannot ensure that each of the people labelling the data used the same logic (or indeed whether they possess the necessary anatomical knowledge). In some cases, those who publish the resulting model openly acknowledge that they were (understandably) not able to manually check the labelling results for all images, due to the volume of data. This can result in a conflict: we input many different images of a body part to the model, but this body part has been labelled in different ways by different people, making it difficult for the model to learn a reliable construct of what that body part looks like (see Table 2 for labelling recommendations).

In addition to OpenPose, many other open source pose estimation models have been trained using open datasets and crowdsourced labels. For example, in Fig. 3, sample images from the commonly used MPII dataset (Andriluka et al., 2014) are shown along with the accompanying crowdsourced labels. In the majority of cases, the labels clearly do not correspond with anatomical landmarks (e.g. the knees in Fig. 3B and the hips in Fig. 3C), which would likely influence the resulting joint angles. Moreover, body parts are labelled even when they are occluded, which in a 2D image usually results in the label being placed on the wrong side of the body (several examples in Fig. 3A). Markers are also placed on different aspects of the same body part (e.g. the ankles in Fig. 3B), which likely makes it more difficult for a neural network to learn appropriate features.

Clearly, crowdsourcing the label process is not appropriate when the goal is to train a model to accurately detect human anatomical landmarks for kinematic analysis. If models are trained using labels that do not reflect the actual body parts that a biomechanist is interested in, the neural network will not 'learn' to label new images correctly. As an example, Fig. 4 shows the output of processing a single running trial with OpenPose (for resulting video see supplementary material), one of the best-known pose estimation algorithms, as well as with manual analysis performed by the author. The OpenPose marker placements for a given body part often exhibit so-called "jitter" between frames (e.g. the hip markers), which is probably at least partly due to the conflicting labelling logic and/or use of multiple labellers mentioned above. The algorithm also sometimes mislabels the right and left limbs. Both of these issues are characteristic of an algorithm that has not robustly learnt to identify specific body parts, and they have important consequences for calculating variables such as body seg-

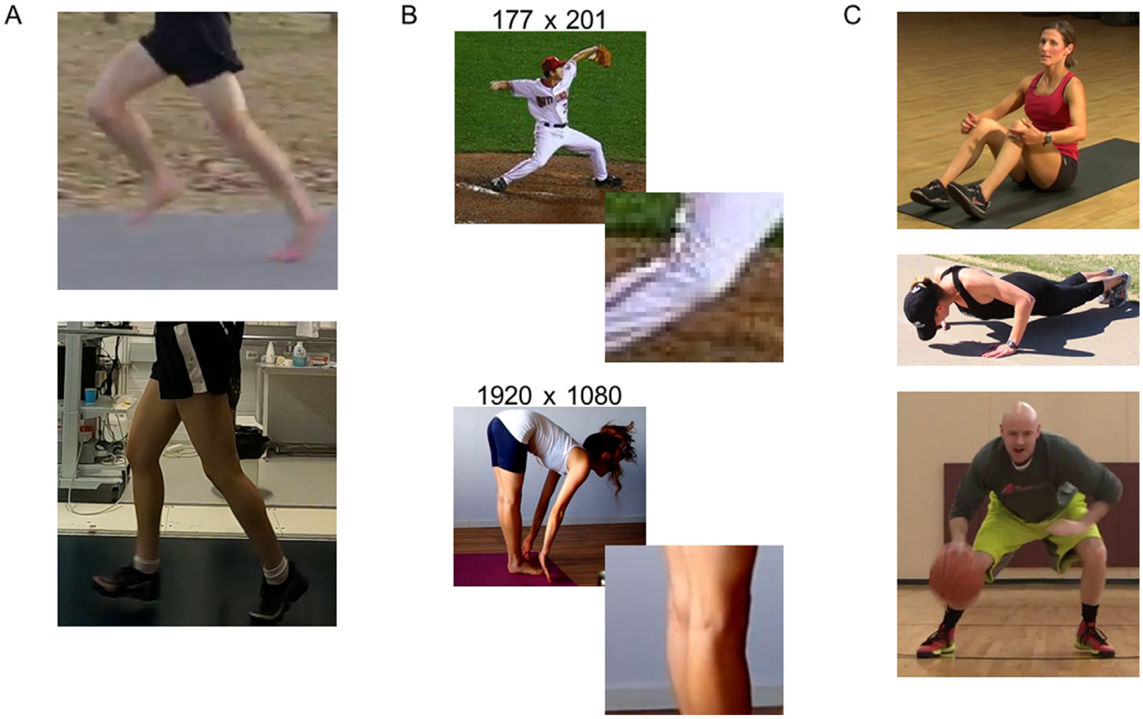

Fig. 2. Challenges associated with labelling 2D images. A: Image blur. The top image is blurry because of a slow shutter speed (1/15). In this case, it would be difficult to accurately label the feet and ankles. In the bottom image, an individual is shown running on a treadmill. This image was extracted from a video sampled at 60 Hz with a shutter speed of 1/250, and the blurring effect is much less evident. B: Image resolution. In the upper image, which is low resolution (177 × 201 pixels; from the Leeds Sports dataset; Johnson and Everingham, 2010), zooming in on a smaller region can make it difficult to accurately identify anatomical landmarks because of pixellation (inset). In the lower image (1920 × 1080 pixels; from the MPII dataset; Andriluka et al., 2014), pixellation is much less evident. However, in both cases the challenge of accurately identifying the joint centre without being able to physically palpate remains a challenge. C: Occlusion. When using a single camera, body parts are often not visible in the resulting 2D images. Occlusion can be in the form of one body part blocking another (top; hand blocking knee), or simply because the target body part is blocked by the body itself (middle; right side not visible) or by an implement (bottom; right knee blocked by ball). All images in C are from MPII (Andriluka et al., 2014).

ment lengths and joint angles (Fig. 4). Based on the above-mentioned limitations and my own experience, some recommendations for labelling are given in Table 2 (see also Mathis et al., 2020).

The issues outlined above lead us to a quandary: we need lots of training examples to produce robust deep neural networks. On the other hand, such large volumes of data cannot be feasibly labelled by a single person, and yet using multiple labellers can affect model accuracy. One possible solution lies in the use of an approach called transfer learning.

4. Transfer learning

Transfer learning involves modifying a model that has been trained to perform a certain task, say, identifying vehicles in images, in such a way that it retains some of the learned features and is then re-trained to identify features in a new set of images, namely those related to our task (Donahue et al., 2013). For example, we might wish to use transfer learning to adapt a model that has been trained to detect different categories of objects in an image (e.g. tables and chairs) so that it can be used to detect human body parts. To do this, we would re-train the existing model by modifying its structure and then showing it a set of labelled images of the anatomical landmarks of interest. Thus, although the detection of tables and chairs is not at all related to our task, there are common features to both tasks that allow for “knowledge” to be transferred between them, hence the name transfer learning (see Johnson et al., 2019 for an example related to human motion analysis).

Because transfer learning involves modifying a model that has already been trained on a large number of images, when we train it to perform a new task, we no longer need huge datasets. DeepLabCut, for example, exploits this concept and can yield models that make accurate predictions despite being trained with just a few hundred training examples (Mathis et al., 2018; Cronin et al., 2019; Papic et al., 2020). It should be noted however that if training images are of low quality (e.g. low resolution, blurring, occlusion), or include movements such as gymnastics or pole vault where the athlete is often upside down, more training images may be required. Nonetheless, it is reasonable to conclude that recent advances have brought us closer to the dream of a truly markerless (and open source) approach that can be used outside of a lab environment, including in natural sports and training settings.

5. What next?

Artificial intelligence has received huge media interest in recent years, and perhaps as a result, people often over-estimate the current state of the art and what can realistically be achieved (Siegel, 2019). It is important to remember that neural networks do not perform magic tricks; they identify mathematical patterns in data. If the dataset used to train a model is small and homogeneous (e.g. only includes data from 2 to 3 individuals), the trained model is very likely to make poor predictions when tested on images of previously unseen people, or even the same people from the training dataset imaged in different settings. In other words, neural networks perform well on the specific task for which they were trained. To produce robust models, the training dataset should include examples of a wide range of human poses, environments, clothing, lighting etc. As with any scientific tool, neural networks have their limitations, but when used appropriately they have the potential to revolutionise the way kinematic analyses are per-

Table 2
Recommendations for working with deep learning pose estimation methods.

Problem	Possible solution(s)	Comments
Blurry training images	Use higher shutter speed (and potentially framerate)	Image blur is relevant during both the training and testing phases and may need to be considered when choosing camera type and settings
Pixellated training images	When aiming to track fast movements (e.g. running), use at least FHD resolution (1920 × 1080) when collecting training data	Higher image resolution makes accurate labelling easier. Images (and labels) can be downscaled prior to neural network training, helping to decrease the memory requirements of training
Inconsistent model performance due to multiple labellers of training images	Agree upon anatomical landmarks in advance; check group agreement on (at least) a subset of images to ensure training data are reliable	With transfer learning approaches (see text), it is usually feasible for multiple individuals to label all training data
Occluded body part	Only label body landmarks that are clearly visible (Papic et al., 2020)	A smaller number of accurate labels is superior to a larger number of inaccurate labels. Poor or inconsistent labelling distorts the model's "understanding" of what a particular part should look like
Inconsistent labelling	Label a given body part consistently, e.g. always the lateral side of the part if visible. If both sides of a joint need to be labelled, label them separately	Generic labelling of a body part (e.g. labelling any visible part of the ankle) necessitates a larger volume of training data for the model to detect that part consistently
Gaps in data labelled by a trained model	Fill gaps using spatio-temporal filtering (Karashchuk, 2020; Papic et al., 2020)	Simpler filters (e.g. median) may suffice when labels are only missing for 1–2 consecutive frames

formed. Although the accuracy of these methods compared to ground truth is still somewhat unclear (D'Antonio et al., 2020; Nakano et al., 2020), this is a very active research area. In fact, it is likely that in the best case, the accuracy of such systems is already sufficient for them to be useful in certain settings, such as for giving rapid feedback to athletes about training performance, or to assist in clinical monitoring of gait disturbances. Ultimately, the biomechanics community will also need to decide how accurate is accurate enough.

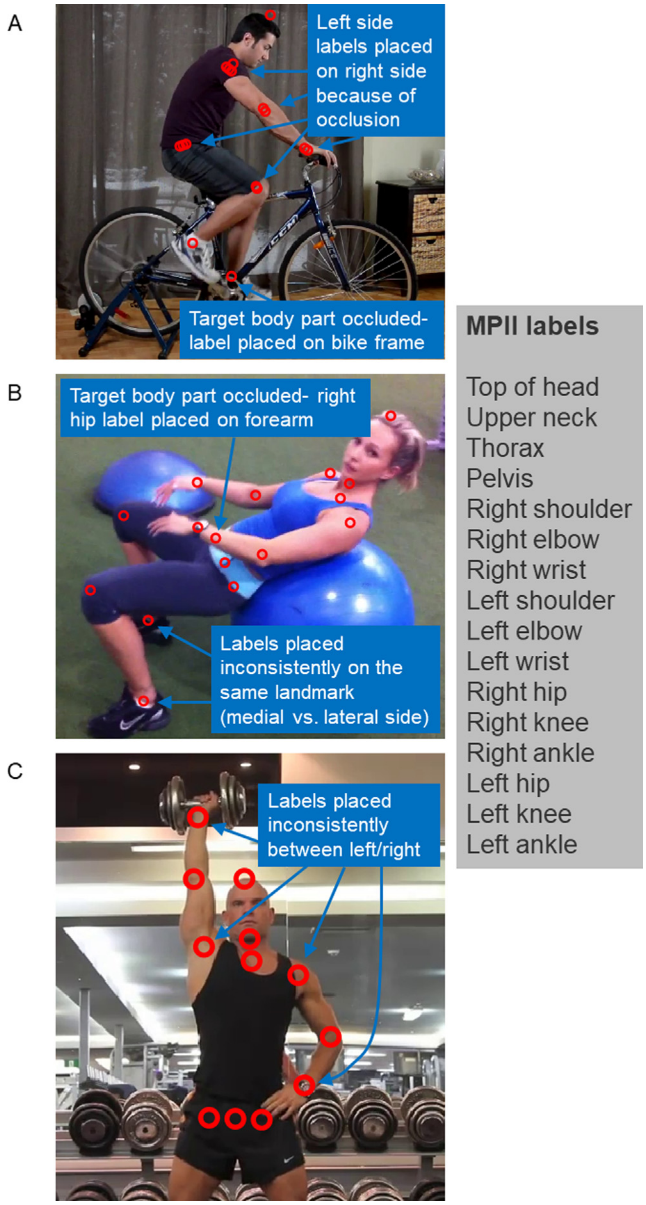
A
Left side labels placed on right side because of occlusion
Target body part occluded- label placed on bike frame
B
Target body part occluded- right hip label placed on forearm
Labels placed inconsistently on the same landmark (medial vs. lateral side)
C
Labels placed inconsistently between left/right
MPII labels
Top of head
Upper neck
Thorax
Pelvis
Right shoulder
Right elbow
Right wrist
Left shoulder
Left elbow
Left wrist
Right hip
Right knee
Right ankle
Left hip
Left knee
Left ankle

Fig. 3. Examples of mislabelled body parts in the MPII dataset (crowdsourced labels are shown with red circles). A: Left-side body part labels are placed on the right side because the left side of the body is not visible. B: Markers for the left and right limb are placed inconsistently, e.g. lateral versus medial side (knees, elbows, ankles). Note also that the right hip marker is placed where the hip is presumed to be, but the label itself is on the left forearm. C: Similar issues to A and B, especially at the shoulders and wrists. Note that the hip markers do not correspond with the greater trochanter location commonly used in optical motion analysis. Inset: the labelling scheme used for the MPII dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Existing techniques rely on supervised learning to detect body landmarks, but there are several exciting new avenues that may eventually help to improve accuracy further. For example, self-supervised learning, as the name suggests, removes the need to manually label training data and instead relies on cues (or "labels") that can be extracted automatically from the data itself (Table 1). This could theoretically allow very large, diverse datasets to be used to train accurate models capable of 2D or even 3D analysis. Promising advances in this area have already been made (e.g. Kocabas et al., 2019; Kundu et al., 2020), although these methods have not yet been applied to the specific task of kinematic analysis, which requires not just the accurate detection of body landmarks, but additional post-processing to yield joint angles. Other neural network approaches such as graph-based methods allow information (or 'knowledge') to be encoded into an algorithm (e.g. Ge et al., 2019). If these techniques could be successfully leveraged, neural networks might learn to make fewer mistakes. Moreover, they would be better equipped to deal with some of the constraints of human movement, such as the fixed connections between body parts, and realistic frame-to-frame changes in joint range of motion or movement velocity.

To date, few comparisons have been performed between markerless and marker-based methods, partly because it is challenging to do reliably. However, OpenPose can predict marker locations that are often within 1–3 cm of the actual anatomical landmarks according to optical systems (Nakano et al., 2020). I believe that this magnitude of error can be improved upon using models tailored to the require-

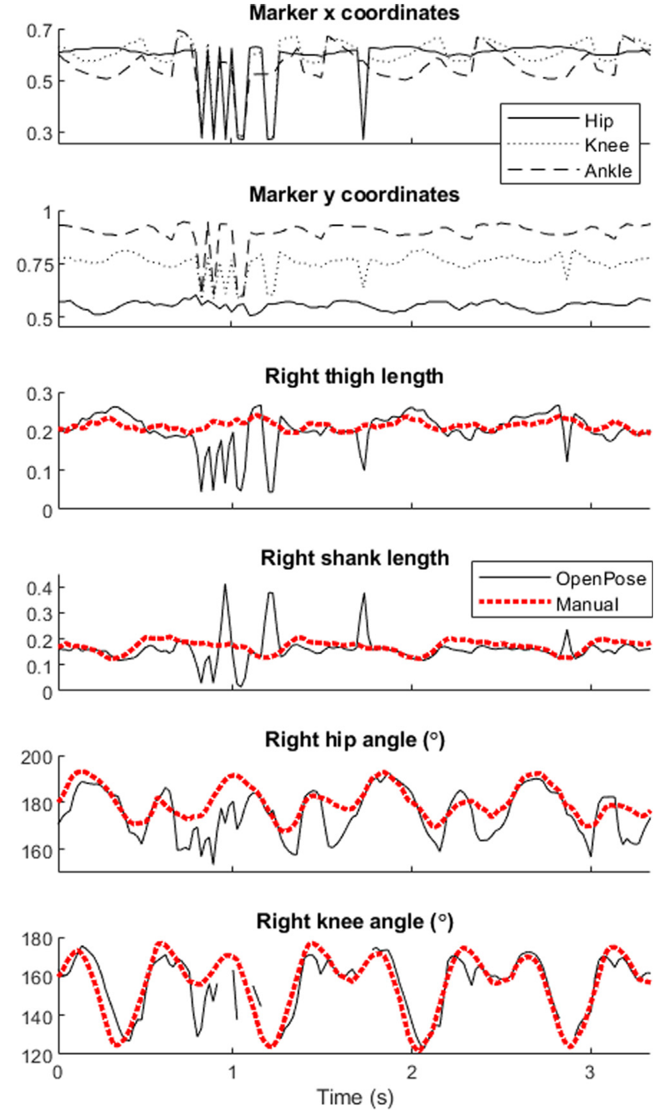

Fig. 4. Sagittal plane analysis of a segment of treadmill running, sampled at 60 Hz with a GoPro HERO8 camera. Note that the x,y coordinates (A and B) often show large, rapid changes in location between frames (e.g. ankle and knee x-coordinates). Segment lengths computed using the distance formula (C and D) also show large, non-physiological deviations over time, which are possible because OpenPose does not constrain segment lengths. Joint angles (E and F) were computed using x,y coordinates predicted by OpenPose (solid black traces). For segment lengths and joint angles, results of the author's manual analysis of the same video are also shown (dotted red traces). This figure clearly shows that inaccurate detection of anatomical landmarks leads to inaccurate joint angle estimates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ments of biomechanical analysis. Given that the current gold standard optical systems and manual digitisation also include inherent limitations (e.g. movement of skin and markers relative to the underlying anatomical landmark), if we reach a state where marker-based and markerless methods yield results within a few mm of each other, markerless motion analysis could truly be a feasible option for human movement scientists, both in and outside of the lab.

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.

Acknowledgements

I am grateful to the Academy of Finland for funding (decision number: 323473), and to Roope Uitto and Leonoor Vlaanderen for their assistance. I also thank the groups of Irene Davis (Harvard Medical School) and Benedicte Vanwanseele (KU Leuven) for valuable discussions about this topic.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbiomech.2021.110460.

References

Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B., 2014. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. CVPR, 3686–3693. https://doi.org/10.1109/CVPR.2014.471.

Arac, A., Zhao, P., Dobkin, B.H., Carmichael, S.T., Golshani, P., 2019. DeepBehavior: A Deep Learning Toolbox for Automated Analysis of Animal and Human Behavior Imaging Data. Front. Syst. Neurosci. 13, 20. https://doi.org/10.3389/fnsys.2019.00020.

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S.-E., Sheikh, Y.A., 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Trans. Pattern Anal. Mach. Intell. 1–1. https://doi.org/10.1109/tpami.2019.2929257.

Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y., 2017. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. CVPR, 1302–1310. https://doi.org/10.1109/CVPR.2017.143.

Colyer, S.L., Evans, M., Cosker, D.P., Salo, A.I.T., 2018. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. Sport. Med. - Open 4, 24. https://doi.org/10.1186/s40798-018-0139-y.

Cronin, N.J., Rantalainen, T., Ahtiainen, J.P., Hynynen, E., Waller, B., 2019. Markerless 2D kinematic analysis of underwater running: A deep learning approach. J. Biomech. 87, 75–82. https://doi.org/10.1016/j.jbiomech.2019.02.021.

Cunningham, P., Cord, M., Delany, S.J., 2008. Supervised learning. In: Cognitive Technologies. Springer Verlag, pp. 21–49.

D'Antonio, E., Taborri, J., Palermo, E., Rossi, S., Patane, F., 2020. A markerless system for gait analysis based on OpenPose library. I2MTC 2020 - International Instrumentation and Measurement Technology Conference, Proceedings. Institute of Electrical and Electronics Engineers Inc..

Davidson, P., Virekunnas, H., Sharma, D., Piché, R., Cronin, N., 2019. Continuous Analysis of Running Mechanics by Means of an Integrated INS/GPS Device. Sensors 19, 1480. https://doi.org/10.3390/s19061480.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database. CVPR, 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2013. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. 31st Int. Conf. Mach. Learn. ICML 2014, 2, pp. 988–996.

Ge, L., Ren, Z., Li, Y., Xue, Z., Wang, Y., Cai, J., Yuan, J., 2019. 3D Hand Shape and Pose Estimation from a Single RGB Image. CVPR, 10833–10842.

Graving, J.M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B.R., Couzin, I.D., 2019. Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. Elife 8. https://doi.org/10.7554/eLife.47994.

Johnson, S, Everingham, M, 2010. Clustered pose and nonlinear appearance models for human pose estimation. Proceedings of the British Machine Vision Conference. https://doi.org/10.5244/C.24.12.

Johnson, W.R., Mian, A., Lloyd, D.G., Alderson, J.A., 2019. On-field player workload exposure and knee injury risk monitoring via deep learning. J. Biomech. 93, 185–193. https://doi.org/10.1016/j.jbiomech.2019.07.002.

Kane, G., Lopes, G., Saunders, J.L., Mathis, A., Mathis, M.W., 2020. Real-time, low-latency closed-loop feedback using markerless posture tracking. bioRxiv 2020.08.04.236422. https://doi.org/10.1101/2020.08.04.236422

Karashchuk, P, et al., 2020. Anipose: a toolkit for robust markerless 3D pose estimation. bioRxiv. https://doi.org/10.1101/2020.05.26.117325.

Kidziński, Ł., Yang, B., Hicks, J.L., Rajagopal, A., Delp, S.L., Schwartz, M.H., 2020. Deep neural networks enable quantitative movement analysis using single-camera videos. Nat. Commun. 11, 1–10. https://doi.org/10.1038/s41467-020-17807-z.

Kocabas, M., Karagoz, S., Akbas, E., 2019. Self-Supervised Learning of 3D Human Pose using Multi-view Geometry. CVPR, 1077–1086.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Int. Conf. Neural Inf. Process. Syst., pp 1097–1105. https://doi.org/10.1145/3065386.

Kundu, J.N., Seth, S., Jampani, V., Rakesh, M., Babu, R.V., Chakraborty, A., 2020. Self-Supervised 3D Human Pose Estimation via Part Guided Novel Image Synthesis. ArXiv, 6151–6161.

Liu, T., Inoue, Y., Shibata, K., 2010. A Wearable Ground Reaction Force Sensor System and Its Application to the Measurement of Extrinsic Gait Variability. Sensors 10, 10240–10255. https://doi.org/10.3390/s101110240.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., Bethge, M., 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. 21, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y.

Mathis, A., Schneider, S., Lauer, J., Mathis, M.W., 2020. A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives. Neuron 108, 44–65. https://doi.org/10.1016/j.neuron.2020.09.017.

Mundt, M., Thomsen, W., Witter, T., Koeppe, A., David, S., Bamer, F., Potthast, W., Markert, B., 2020. Prediction of lower limb joint angles and moments during gait using artificial neural networks. Med. Biol. Eng. Comput. 58, 211–225. https://doi.org/10.1007/s11517-019-02061-3.

Nakano, N., Sakura, T., Ueda, K., Omura, L., Kimura, A., Iino, Y., Fukashiro, S., Yoshioka, S., 2020. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. Front. Sport. Act. Living 2, 50. https://doi.org/10.3389/fspor.2020.00050.

Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., Mathis, M.W., 2019. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. Nat. Protoc. 14, 2152–2176. https://doi.org/10.1101/476531.

Nowak, S., Rüger, S., 2010. How reliable are annotations via crowdsourcing? A study about inter-annotator agreement for multi-label image annotation. In: Proceedings of the International Conference on Multimedia Information Retrieval, p. 557. https://doi.org/10.1145/1743384.1743478

Papic, C., Sanders, R.H., Naemi, R., Elipot, M., Andersen, J., 2020. Improving data acquisition speed and accuracy in sport using neural networks. J. Sports Sci. https://doi.org/10.1080/02640414.2020.1832735.

Seethapathi, N., Wang, S., Saluja, R., Blohm, G., Kording, K.P., 2019. Movement Science Needs Different Pose Tracking Algorithms. ArXiv, p. arXiv:1907.10226.

Siegel, E., 2019. The Media's Coverage of AI is Bogus [WWW Document]. Sci. Am. URL https://blogs.scientificamerican.com/observations/the-medias-coverage-of-ai-is-bogus/

Tome, D., Toso, M., Agapito, L., Russell, C., 2018. Rethinking Pose in 3D: Multi-stage Refinement and Recovery for Markerless Motion Capture. Proc. - 2018 Int. Conf. 3D Vision, 3DV, 2018, pp. 474–483.

Zimmermann, T., Taetz, B., Bleser, G., 2018. IMU-to-Segment Assignment and Orientation Alignment for the Lower Body Using Deep Learning. Sensors 18, 302. https://doi.org/10.3390/s18010302.