Journal of Biomechanics 87 (2019) 75–82

Contents lists available at ScienceDirect

Journal of Biomechanics

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

Markerless 2D kinematic analysis of underwater running: A deep learning approach

Neil J. Cronin [a,*], Timo Rantalainen [a], Juha P. Ahtiainen [a], Esa Hynynen [b], Ben Waller [a,c]

[a] Faculty of Sport and Health Sciences, University of Jyväskylä, Finland
[b] KIHU- Research Institute for Olympic Sports, Jyväskylä, Finland
[c] Physical Activity, Physical Education, Sport and Health Research Centre (PAPESH), Sports Science Department, School of Science and Engineering, Reykjavik University, Reykjavik, Iceland

ARTICLE INFO

Article history:
Accepted 22 February 2019

Keywords:
Deep water running
Kinematics
Deep learning
Artificial intelligence
Motion analysis

ABSTRACT

Kinematic analysis is often performed with a camera system combined with reflective markers placed over bony landmarks. This method is restrictive (and often expensive), and limits the ability to perform analyses outside of the lab. In the present study, we used a markerless deep learning-based method to perform 2D kinematic analysis of deep water running, a task that poses several challenges to image processing methods. A single GoPro camera recorded sagittal plane lower limb motion. A deep neural network was trained using data from 17 individuals, and then used to predict the locations of markers that approximated joint centres. We found that 300–400 labelled images were sufficient to train the network to be able to position joint markers with an accuracy similar to that of a human labeler (mean difference < 3 pixels, around 1 cm). This level of accuracy is sufficient for many 2D applications, such as sports biomechanics, coaching/training, and rehabilitation. The method was sensitive enough to differentiate between closely-spaced running cadences (45–85 strides per minute in increments of 5). We also found high test–retest reliability of mean stride data, with between-session correlation coefficients of 0.90–0.97. Our approach represents a low-cost, adaptable solution for kinematic analysis, and could easily be modified for use in other movements and settings. Using additional cameras, this approach could also be used to perform 3D analyses. The method presented here may have broad applications in different fields, for example by enabling markerless motion analysis to be performed during rehabilitation, training or even competition environments.

1. Introduction

Kinematic analysis is used to characterise changes in joint angles during human movement. This information can be combined with other sources, e.g. force data, to build a more complete picture of how a movement is performed (Winter, 1991), and thus has important implications for various fields such as sports biomechanics, injury risk assessment and rehabilitation (see Colyer et al. 2018 for a review). Kinematic analysis is often performed with a camera system combined with a set of reflective markers placed over bony landmarks, allowing a digital model of the moving person to be reconstructed (van der Kruk and Reijne, 2018). However, the use of reflective markers can restrict the settings in which data can realistically be collected, and many existing camera-based methods still rely on expensive hardware and software. Moreover, in an aquatic environment, the use of markers is impractical because they impede normal movement and are prone to significant motion artifact.

Recently, several attempts have been made to develop markerless methods, which in theory could be used outside of the laboratory and allow movement to be analysed in more natural, unconstrained conditions (see Drory, Li, and Hartley 2017 for a comprehensive overview). In particular, methods that rely on artificial intelligence have demonstrated promising results (see Colyer et al., 2018 for review), and have the potential to revolutionise the way movement analysis is performed due to their powerful ability to 'learn' patterns in data. In the present study, we used DeepLabCut (Insafutdinov et al., 2016; Mathis et al., 2018; Pishchulin and Insafutdinov, 2015) to track the locations of (approximated) lower limb joint centres and used this information to perform 2D kinematic analysis of deep water running, a task that poses several

* Corresponding author at: University of Jyvaskyla, Neuromuscular Research Center, Faculty of Sport and Health Sciences, P. O. Box 35, FI-40014, University of Jyvaskyla, Finland.
E-mail address: neil.j.cronin@jyu.fi (N.J. Cronin).

https://doi.org/10.1016/j.jbiomech.2019.02.021
0021-9290/

challenges to image processing methods, such as poor contrast and changes in light intensity. DeepLabCut is an open-source method that combines a residual neural network (ResNet-50) pretrained on ImageNet with deep convolutional and deconvolutional neural network layers (Insafutdinov et al., 2016) to predict the 'learned' locations of individual points in an image using feature detectors (He et al. 2015). The network 'learns' marker locations by being trained on labeled data, which consists of individual images accompanied by a human-defined label of the 'correct' marker location. During training, the weights are adjusted iteratively so that for each image, the network assigns high probabilities to target marker locations and low probabilities to all other regions (Mathis et al., 2018). Training thus allows the network to 'learn' feature detectors for each user-defined marker, rather than relying on hard-coded, pre-defined features.

In this study we demonstrate that a modified version of the DeepLabCut method can be used for accurate 2D kinematic analysis of deep water running filmed using a single GoPro camera. We used this method to determine lower limb segment lengths and joint angles, and we present various other parameters that could be useful in motion analysis applications.

2. Methods

2.1. Participants

A total of 21 individuals (age: 24 ± 4 years, height: 177 ± 10 cm, mass 67 ± 9; 13 males and 8 females) volunteered to participate and provided written informed consent. The study was approved by the University's ethics committee, and testing was conducted in accordance with the most recent Helsinki declaration.

2.2. Experimental protocol

Participants performed bouts of deep water running whilst immersed to shoulder level, and were tethered to the edge of the pool by a non-elastic cable attached to a buoyancy aid (Aquawallgym©, Hungary). A single GoPro camera (Hero 3 model) was enclosed in a waterproof case and positioned underwater in the sagittal plane to the participants' left side at a distance of approximately 5 m. A custom-made calibration frame (2 m × 2 m) was used to calibrate the field of view for each participant and test. The camera was then set to record at 60 Hz whilst participants 'ran' at different cadences controlled by a metronome (increased by 5 strides per minutes (spm) from 45 to 90 spm). A deep neural network was trained and then used to predict the locations of several markers that approximated joint centres. Predicted joint coordinates were used to determine lower limb segment lengths and joint angles from the left leg, which was closest to the camera.

2.3. Deep Neural Network

The method used here largely followed the method described by Mathis et al. (Mathis et al., 2018; v1). We first trained the network using 500 images from 17 randomly chosen participants (i.e. 28–30 images per participant), leaving aside data from the remaining 4 participants (see below). The training images were randomly selected using a custom-written script in Matlab (Mathworks, v2016b). These images were cropped (dimensions: 580 × 480 pixels) and then manually labelled, with markers placed on the lateral side of the trunk (approximately mid-way between the shoulder and hip), greater trochanter, lateral femoral condyle, lateral malleolus, and 5th meta-tarsal head. The labelled images were used to train a deep neural network with a 90% training, 10% test split. The ResNet model was initialised with weights trained on ImageNet (He et al., 2015), and the cross-entropy loss between the predicted score-map and the ground truth score-map was minimised using stochastic gradient descent (Insafutdinov et al., 2016). The network was trained for 200,000 iterations using a single Tesla K80 GPU via Microsoft Azure's cloud platform running Python (Python Software Foundation; v.3.5) and Tensorflow (Abadi et al., 2015; v.1.2.1). The training process was repeated with smaller training sets (400, 300, 200 and 100 images respectively), to determine the minimum number of images required to reach satisfactory predictive performance for this task. The number of frames used for training was selected based on previous work using a similar method (Mathis et al., 2018), and for each trained model, frames were randomly assigned to the test or training set.

2.4. Evaluation of deep neural network performance

To compare between joint coordinates labelled by a human and those labelled by the network, pairwise Euclidean distances were computed for each marker location (root mean square error: RMSE). In the results section the RMSE values are shown for individual joints or as the average across all joints, as appropriate. To quantify the evolution of the training error, and to enable training to be resumed later if needed, the Tensorflow weights were stored every 10,000 iterations. As noted above, data from 4 randomly chosen participants were excluded completely from the training set. After training of the neural networks was complete, videos from these 4 participants were evaluated by each neural network model, thereby serving as additional test data. This approach was chosen to enable out of sample predictions that were completely independent of the training process, thus giving some indication of the generalisability of our trained models.

2.5. Determining joint angles and segment lengths

Segment lengths were initially computed in terms of pixels, using the coordinate data of each point exported during the analysis of each video. Segment lengths were calculated based on the distance formula: $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$,

where d = segment length in pixels, and x and y values denote the coordinates of the two points that make up a segment. A scaling factor was calculated for each participant and trial based on the corresponding calibration frame video, and used to scale segment lengths. Joint angles at the hip, knee and ankle were determined using the atan2 function in Matlab. In several cases, the neural network did not attempt to place a marker because the target joint location was blocked by the hand or moved beyond the image field of view. To overcome the effect of missed (and misplaced) markers on the resulting kinematic and segment length data, raw data were first filtered with a median filter (10–20 data points generally yielded good results) followed by a Butterworth 4th order low-pass filter (Fig. 3). In some cases, e.g. when a marker was missing for several consecutive frames, it was necessary to experiment with different filtering procedures.

3. Results

3.1. Deep neural network performance

Using the full training set of 500 labelled images, the mean training error across all images was 1.4 pixels. The mean test error was 2.92 pixels (approximately 1 cm). This model represents the best performance achieved out of all of the tested models. As seen in Fig. 1A, training performance was similar between all of the

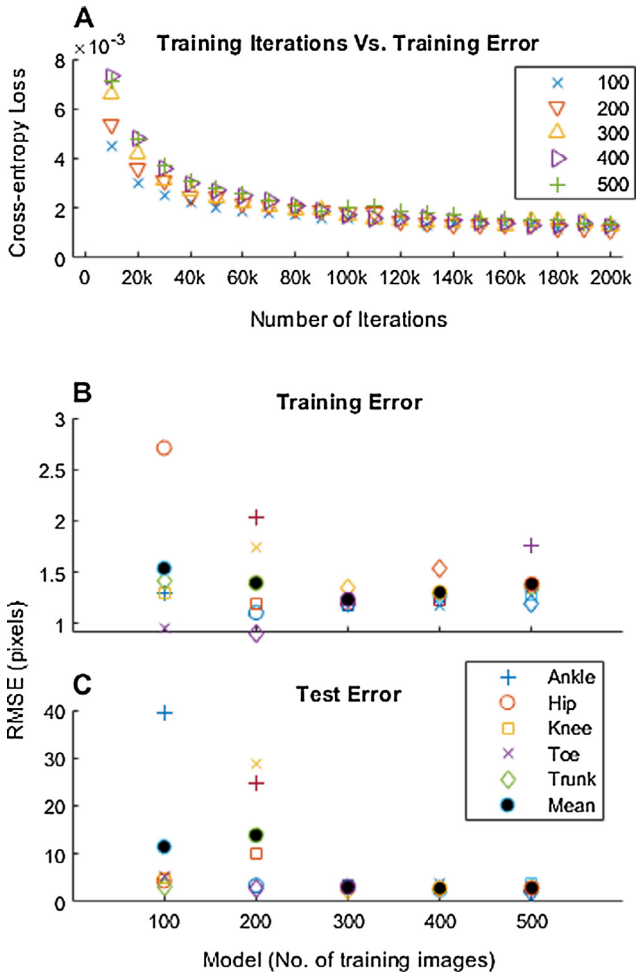


Fig. 1. A: Mean training error as a function of the number of iterations. The legend identifies different models (different numbers of training images). B: Training error of each trained network showing data for individual markers and for the mean of all markers (filled circles). C: Test error presented in the same way as in B.

tested models after 200,000 iterations. Test performance, i.e. how well the network predicts marker locations on images it has not 'seen' during training, was similar between models trained on 300–500 images, suggesting that 300 training images was sufficient for this task. However, test performance clearly decreased with training datasets of 100–200 images, indicating overfitting of these models during the training stage. For all models, training time varied between approximately 9–12 h.

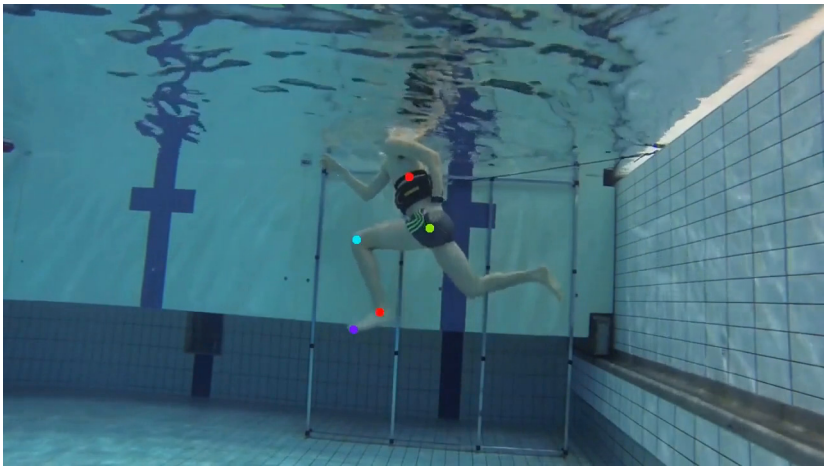
For the better performing models, both training and test errors were largely independent of which marker was being tracked, whereas for the poorer performing models, the disparity between different markers was much larger (Fig. 1B and C). It should be noted that the trunk marker was not placed by the network in around 20% of frames due to the hand blocking the target area. In some images, the 5th metatarsal marker also was not placed because the foot moved beyond the camera field of view. However, these issues did not substantially affect kinematic tracking (Figs. 3 and 4).

Fig. 2 shows some examples of the same images labelled by the 100 and 500 models. In some cases, the models make similar predictions, compared to each other and to a human labeller (e.g. middle image in Fig. 2). In other images, the 100 model consistently makes larger errors, and in the most extreme cases, identifies an ostensibly correct location but on the wrong limb (left image in Fig. 2), which largely accounts for the bigger test errors of the poorer performing models (for further model comparisons see Supplementary Video 1). Segment length calculations yielded consistent traces across consecutive stride cycles, particularly for models trained on more images. Segment lengths varied somewhat throughout a stride (see Fig. S1), due to minor fluctuations in marker locations, as well as the inevitable 3D rotation of the lower limb that cannot be quantified with this method.

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jbiomech.2019.02.021.

3.2. Kinematics of deep water running

Using this method we obtained consistent joint angle traces over several consecutive stride cycles. Fig. 3 shows examples of data computed from three different 10 s videos obtained from different individuals whose data were not seen by the neural network during training. These videos were processed entirely by a trained neural network, and did not require a human labeler at any stage (see also Supplementary Videos 2–4).



Supplementary Video 1.

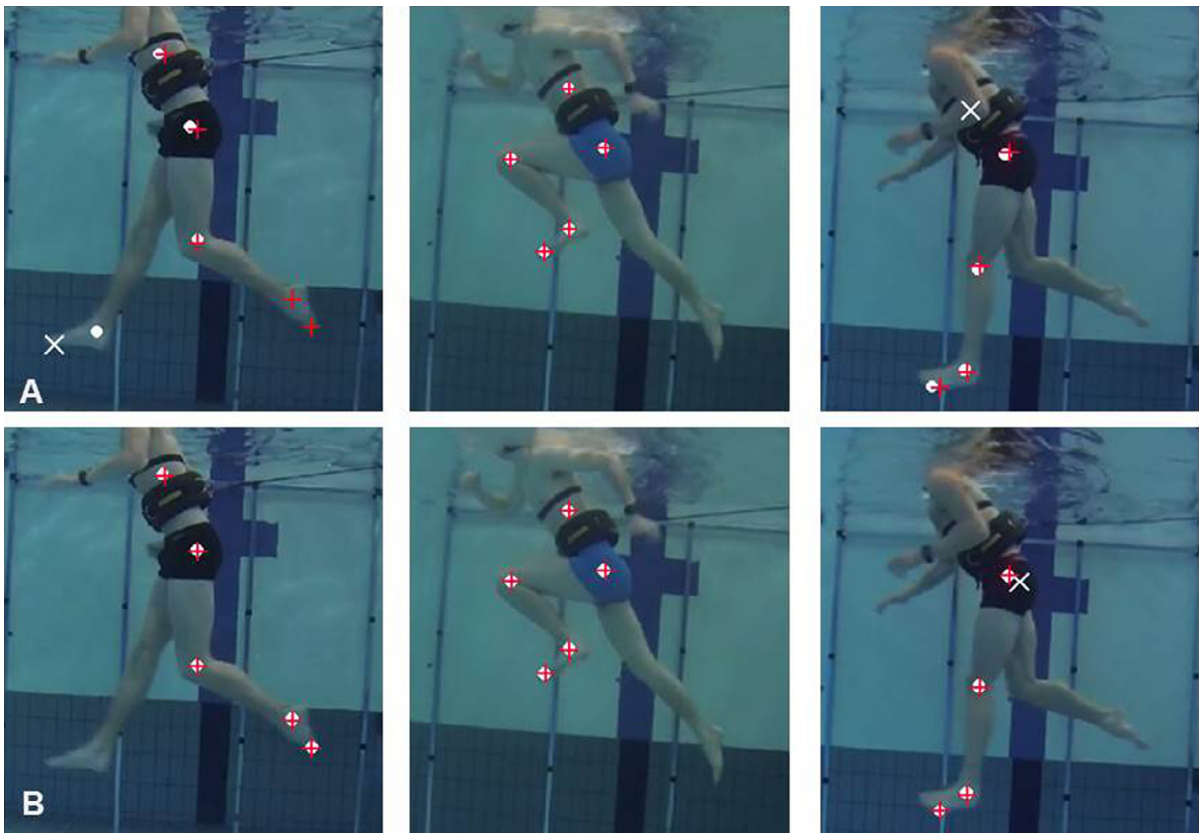


Fig. 2. Examples of labelled images from the network trained on 100 images (A), and the corresponding labels for the network trained on 500 images. Human labels are denoted by red '+' symbols. Network-applied labels (white) where the network confidence score is greater than 0.1 are shown by filled circles, and network-applied labels with a confidence score below 0.1 are shown as 'X'. Note that in A (left image), the network identified the wrong limb when placing the ankle and toe markers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

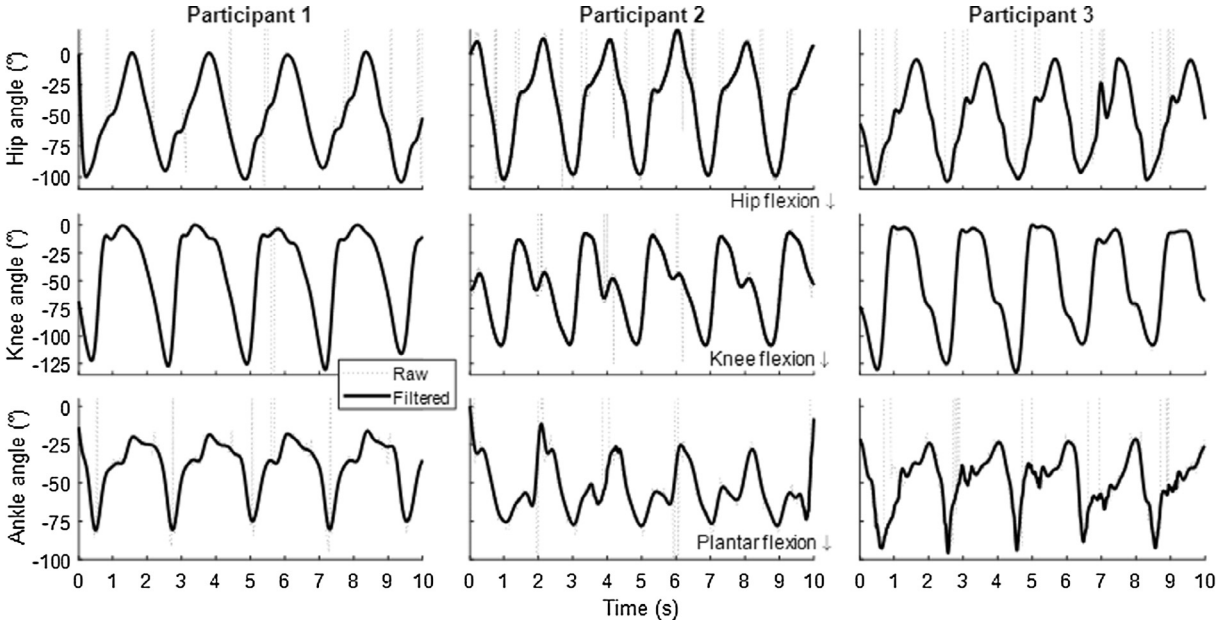
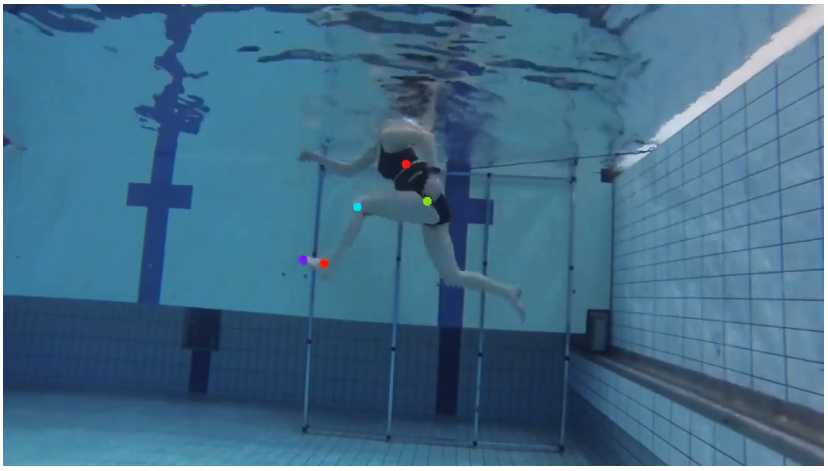


Fig. 3. Raw (dotted lines) and filtered (solid lines) hip, knee and ankle angles for three trials from three different participants whose data were not seen by the neural network during training. These results were computed by the model trained on 300 images.

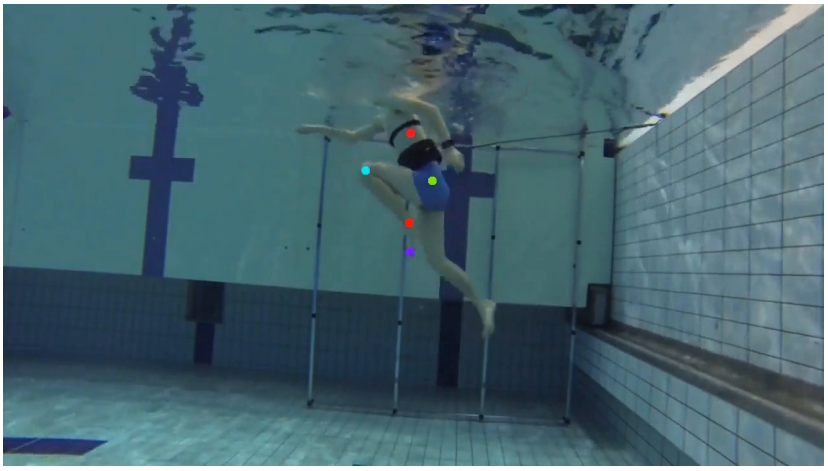
Based on visual identification from the videos, it was possible to approximate the start of individual stride cycles. Fig. 4 shows the results of this segmentation for a single 20 s trial from a participant whose data were not seen by the neural network during training.

For the data shown in Fig. 4, the mean range of motion at the hip joint was 96.7 ± 5.4°. Corresponding values for the knee and ankle joint were 124.0 ± 8.2° and 59.0 ± 6.3° respectively. Similar range of motion values were also obtained for the 3 participants'

Supplementary Video 2. Video from a single participant whose data were not seen by the neural network during training. First the raw video is shown, followed by the markers positioned by the trained neural network (trained on 500 images). Stick figures are then overlaid on the joint marker centres. The red lines that are occasionally visible indicate where markers were briefly misplaced, and the white lines are fitted to the filtered marker coordinates.



Supplementary Video 3. Similar to Supplementary video 2, but using data from a different participant.



Supplementary Video 4. Similar to Supplementary videos 2 and 3 but using data from a different participant.

data in Fig. 3 (hip: 102.2–121.7°; knee: 102.0–133.0°; ankle: 67.2–78.2°). To demonstrate some additional applications of our method, we used it to examine kinematics at a range of different cadences, to ensure that the method was sufficiently robust to small changes in movement velocity and the resulting kinematics. These results are shown in Fig. 5, demonstrating that small changes in cadence of 5 spm can be distinguished reliably based on the kinematic traces.

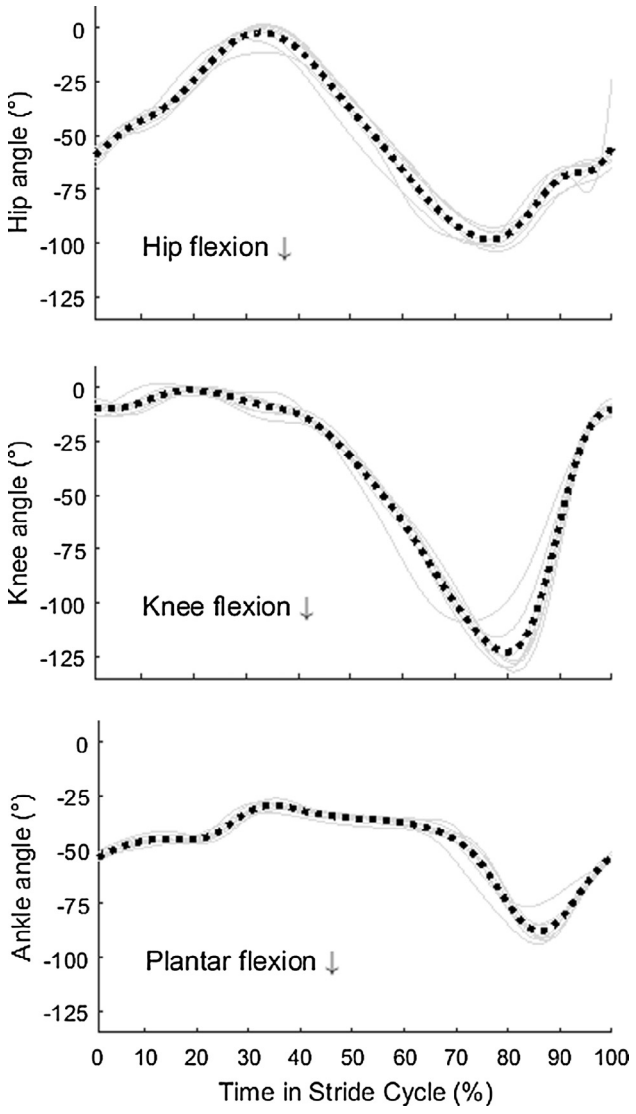


Fig. 4. Filtered data from a 20 s video from one participant separated into 8 individual 'strides', where the start of the cycle coincides with the left leg at approximately bottom dead centre, loosely equivalent to the start of the contact phase in overground running. Individual steps are shown in grey (solid lines) and the mean of all traces is shown in black (dotted lines).

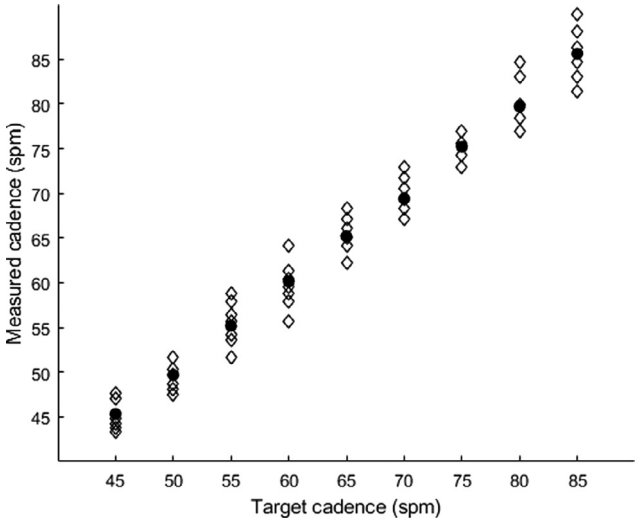


Fig. 5. Data from one participant running at 9 cadences. Target cadence was controlled by a metronome, and measured cadence represents the value determined from joint kinematics computed with our method. Individual strides over ~30 s per condition are represented by diamonds, and filled circles represent the mean.

We also performed test-retest comparisons on data collected from a single individual one week apart (Fig. 6). Fig. 6 shows a segment of data (~30 s), as well as individual strides from each session. Based on the mean strides, the range of motion values for tests 1 and 2 were 107.9° and 102.1° (hip), 115.2 ° and 121.9° (knee), and 35.5° and 33.8° (ankle) respectively. Corresponding mean differentials of these traces were 0.004 and 0.011 (hip), −0.014 and −0.011 (knee), and −0.062 and 0.003 (ankle). Correlation coefficients computed on the pairwise mean stride data showed values of 0.97, 0.90 and 0.93 using the raw data, and 0.93, 0.78 and 0.79 when computed on the differential of the mean stride data, for the hip, knee and ankle respectively.

4. Discussion

In this paper we demonstrate the ability to perform markerless 2D kinematic tracking using a deep residual neural network trained on human-labelled data. Our results show that 300–400 labelled images were sufficient to train the network to be able to position joint markers with an accuracy similar to that of a human labeler (with a mean difference of around 1 cm). This level of accuracy is sufficient for many 2D applications, such as sports biomechanics and coaching, and rehabilitation/training scenarios. Moreover, it is likely that network performance could be further improved, for example by using a deeper pre-trained network or by modifying model hyperparameters (Mathis et al., 2018). In addition to assessing joint kinematics, we also demonstrate the ability to compute relevant parameters such as joint range of motion and cadence on a stride by stride basis, and show strong test-retest reliability of kinematics measured with this method.

The kinematic results obtained in this study are largely comparable to those of the few previous studies conducted in this area. For example, our joint range of motion results are similar to values reported by Kato et al. (2001) and Kilding et al. (2007). For some participants we observed larger hip range of motion than in the Kato study, but this is likely due to the unconstrained nature of deep water running, compared to running on a treadmill in Kato's study. Similarly, we observed less peak knee flexion than Kilding et al., likely due to differences in deep water running technique (high knees versus our modified cross-country technique). At all joints, the kinematic traces in our study were qualitatively similar to those observed in overground running (e.g. Voloshina and Ferris, 2015).

We applied Deep Learning to a task that is very challenging from an image processing perspective. For example, the light intensity of the image background varied between (and even within) tests due to the fact that data were collected in a public swimming pool. The camera used in the present study had automatic shutter speed, and due to the low amount of light, motion blur was evident in the videos, particularly in the distal portion of the image, which may have contributed to the larger RMSE at the ankle than at other joints for some models (Fig. 1). This could conceivably have increased errors in marker placement by both the human and the neural networks. As light is filtered by the water, the contrast of the videos also seemed to be low. With these constraints, image quality was arguably quite low, further exacerbated by isolating and cropping individual video frames during training. It seems likely that using a more advanced camera could have improved overall image quality and thereby minimised

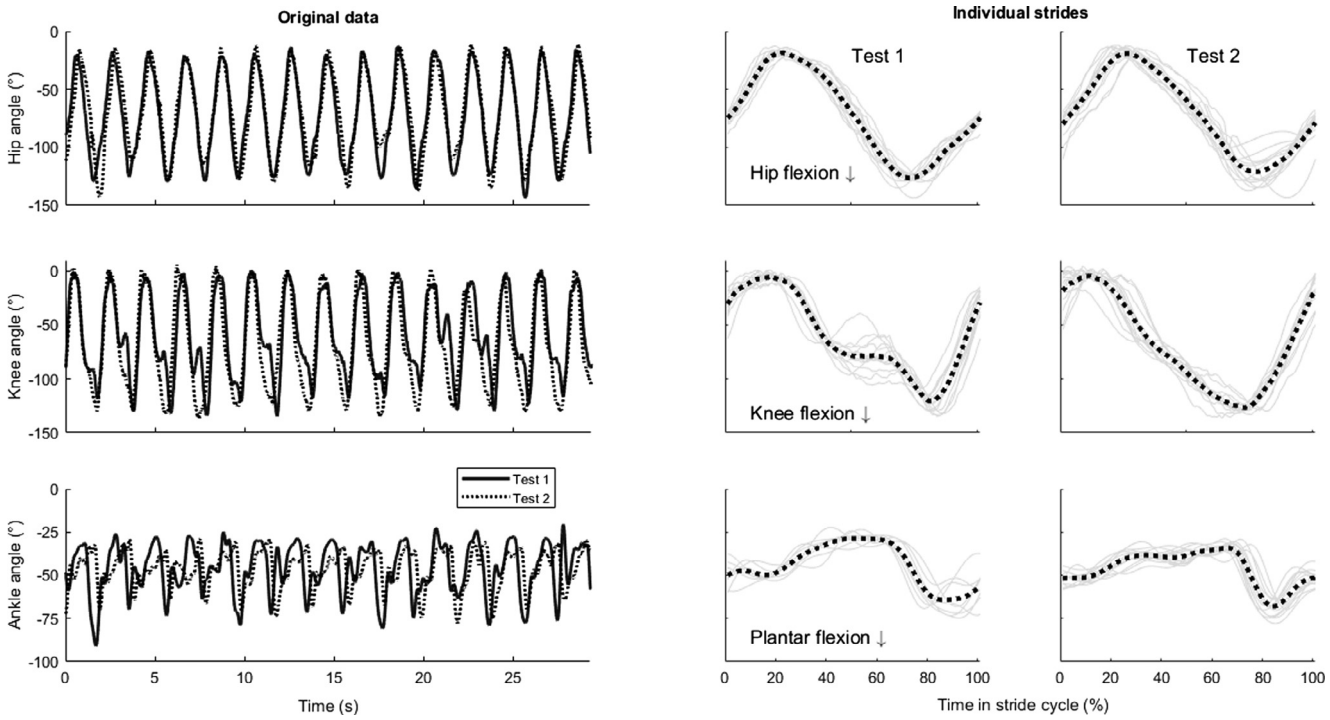


Fig. 6. Test-retest data from an individual running at 60 spm on two different days. The left panels show continuous traces of hip, knee and ankle angles. The right panels show individual strides (grey traces) and the mean stride (dotted black lines) from the data in the panels on the left for each test session.

tracking errors (for human and neural network labelling). However, we see this as a strength of the present approach, since it highlights the robustness of the method in spite of the factors mentioned above.

Aside from issues related to filming underwater, we also encountered difficulties common to gait analysis such as an arm blocking a marker's position. Additional cameras were not necessary to overcome this issue, and a simple filtering procedure combined with a robust deep neural network was sufficient to produce consistent kinematic results. Nonetheless, implementing this method in 3D may help to reduce the effect of marker occlusion, due to the redundancy provided by additional cameras (see Drory et al., 2017 for a similar approach based on single images). Other difficulties included the occasional placement of a marker on the wrong limb by the neural network. To overcome this issue, other studies have used a neural network to learn spatial relations between markers (e.g. the hip marker is always close to the knee marker) to better inform predictions (Drory et al., 2017), and this approach could have helped to improve accuracy in the present study.

We only compared neural network performance to that of a human labeler, and could not evaluate our method against traditional systems that use reflective markers. However, reflective markers in the image would influence neural network performance during training, with a high risk that the network would simply learn to identify the reflective markers, and subsequently fail when used to predict marker positions in images where reflective markers are not present. Overcoming this issue thus requires an alternative approach. Indeed, it should be noted that camera-based methods are not the only possible solution for kinematic analysis. For example, some studies have used inertial measurement units (Dadashi et al., 2012), with the advantage that cameras are not needed, and so the issue of placing markers is avoided completely. However, wearable sensors do still need to be attached to the person, and in our experience, athletes often prefer to avoid instrumentation completely, even very unobtrusive contemporary wearable sensors. Moreover, some sporting federations do not allow wearables in competition at all, thus necessitating a different approach for field-based motion analysis.

The method used here offers a very low-cost, adaptable solution for simple kinematic analysis. The method only requires a small amount of manual labelling of image frames, and in the best case, this training process only needs to be performed once. The successfully trained network can then be used to label new videos quickly (45 s for a 10 s video on a standard CPU), and near real-time tracking is also possible with GPU support (Nath et al., 2018). Given the challenges associated with imaging deep water running, it is likely that this approach could easily be modified to analyse kinematics in other human movements and measurement settings, simply by re-training the network using a suitable dataset. Moreover, using additional cameras, this approach could be used to perform 3D analyses (Nath et al., 2018). As stated by Colyer et al. (2018), the development of methods aided by artificial intelligence could revolutionise sports biomechanics and rehabilitation by broadening the applications of motion analysis to training or even competition environments.

Acknowledgements

The authors gratefully acknowledge the technical assistance of Markku Ruuskanen during data collection.

Conflict of interest statement

None of the authors have any conflicts of interest to declare, financial or otherwise.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O.,

Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., Research, G., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.

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.

Dadashi, F., Crettenand, F., Millet, G.P., Aminian, K., 2012. Front-crawl instantaneous velocity estimation using a wearable inertial measurement unit. Sensors 12, 12927–12939. https://doi.org/10.3390/s121012927.

Drory, A., Li, H., Hartley, R., 2017. A learning-based markerless approach for full-body kinematics estimation in-natura from a single image. J. Biomech. 55, 1–10. https://doi.org/10.1016/J.JBIOMECH.2017.01.028.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv.

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., Schiele, B., 2016. DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model. arXiv.

Kato, T., Onishi, S., Kitagawa, K., 2001. Kinematical analysis of underwater walking and running. Sport. Med. Train. Rehabil. 10, 165–182. https://doi.org/10.1080/10578310210396.

Kilding, A.E., Scott, M.A., Mullineaux, D.R., 2007. A kinematic comparison of deep water running and overground running in endurance runners. J. Strength Cond. Res. 21, 476. https://doi.org/10.1519/R-17975.1.

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.

Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., Mathis, M.W., 2018. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. bioRxiv 476531. http://doi.org/10.1101/476531

Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P., Schiele, B., 2015. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. arXiv.

van der Kruk, E., Reijne, M.M., 2018. Accuracy of human motion capture systems for sport applications; state-of-the-art review. Eur. J. Sport Sci. 18, 806–819. https://doi.org/10.1080/17461391.2018.1463397.

Voloshina, A.S., Ferris, D.P., 2015. Biomechanics and energetics of running on uneven terrain. J. Exp. Biol. 218, 711–719. https://doi.org/10.1242/jeb.106518.

Winter, D.A., 1991. The biomechanics and motor control of human gait : normal, elderly and pathological. University of Waterloo Press.