The use of transfer learning to take estimation of running ground reaction forces via wearables from the treadmill to overground

Joshua Carter[1], Xi Chen[2], Dario Cazzola[1], Grant Trewartha[3], Ezio Preatoni[1]
[1]Department for Health, University of Bath, Bath, UK
[2]Department of Computer Science, University of Bath, Bath, UK
[3]School of Health and Life Sciences, Teesside University, Middlesborough, UK
Email: jc2369@bath.ac.uk

Summary

This study found that a transfer learning approach using ground reaction force (GRF) data from treadmill running helped improve GRF estimation accuracy of a deep learning model during overground running. Leveraging a treadmill dataset 70 times larger than the overground dataset enabled the model to learn generalisable features that could be adapted to the target use case, resulting in more accurate overground GRF predictions than training on the smaller dataset alone.

Introduction

The use of wearable sensors and machine learning for estimating lower limb kinetics during running is becoming increasingly common [1]. However, deep learning models require large, diverse datasets for training, which can be challenging to obtain in biomechanics. Transfer learning offers a solution by leveraging models trained on larger datasets in parallel domains to improve performance on data-scarce tasks. This study aimed to evaluate whether transfer learning could enhance ground reaction force (GRF) estimation during overground running by utilizing a larger dataset from instrumented treadmill running.

Methods

Overground dataset: All 15 participants (6 female) in this study were first fitted with a NURVV Run system in their typical running shoe. The wearable system collected pressure data from 16 force sensitive resistors under each foot (1000 Hz) and an inertial measurement unit (IMU) (1125 Hz) attached to the lateral aspect of the left shoe. Ground truth GRF data was collected from 4 floor embedded force plates. A single reflective marker placed on the waistband was tracked by a 10 camera Qualisys motion capture system, as a proxy measure of running speed. Participants completed three 2-minute running bouts (self-selected slow, medium, fast conditions), repeatedly looping across the four force plates. The marker and force data were synchronized at collection through Qualisys software, whereas the NURVV system was synchronized with the lab system post-hoc by aligning pressure and force signals. The force data was low pass filtered with a 20 Hz cut off. A 20 N threshold was used to identify initial contact and toe off timings. These timings were then used to segment all signals, which were then time normalized to 101 data points.

Treadmill dataset: This larger dataset (Figure 1) was collected with a similar setup [2], besides force data, which came from an instrumented treadmill (1000 Hz).

The deep learning architecture used throughout this study, was comprised of a recurrent neural network with a bi-directional Long Short Term Memory (LSTM) component. Independent models were trained for the estimation of vertical GRF (vGRF) and anteroposterior GRF (apGRF). Signal estimation error was evaluated using relative root mean squared error (rRMSE), where the RMSE was calculated between the measured and estimated force signal and then normalized to the range of the gold standard signal. Model performance was evaluated in a leave one subject out fashion.

Results and Discussion

Across both vGRF and apGRF estimation the mixed domain training outperformed either training on the larger Treadmill Dataset or on the smaller but specific Overground data (Figure 1). A single round of training on a combined dataset slightly outperformed the transfer learning approach. After combined training the mean vGRF rRMSE was 6.1% (range: 3.5-10.7%) and the mean apGRF rRMSE was 5.5% (range: 3.2-7.9%).

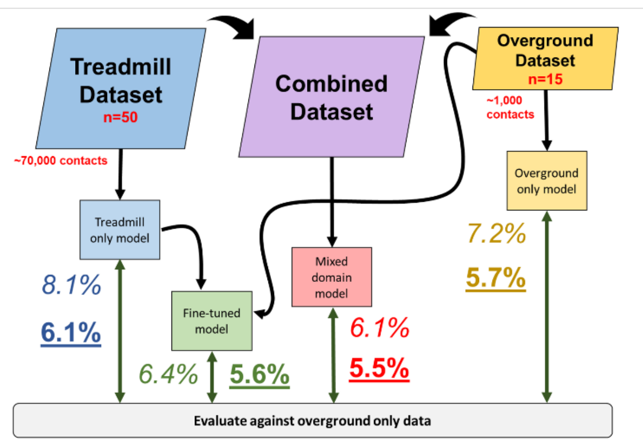

Figure 1: Performance of the different deep learning models, mean relative root mean squared error (rRMSE) for *vGRF* (italics) and **apGRF** (bold and underlined) estimation

Conclusions

The combination of a larger dataset from a similar movement and a smaller dataset specific to the validation condition led to improved estimation accuracy in comparison to the smaller domain specific dataset. This approach could be used to help with the generalization of deep learning models in other biomechanical contexts, where data scarcity is often a challenge, such as different surfaces or movement types.

References

[1] Xiang et al. (2022). *Front. neurorobot*, **16**: 913052.
[2] Carter et al. (2024) *PeerJ*, **12**: 17896