Classifying humans using deep time-series transfer learning : accelerometric gait-cycles to gyroscopic squats

Vinay Prabhu*
UnifyID AI Labs
Redwood City, CA
vinay@unify.id

Stephanie Tietz*
Stanford University
Stanford, CA
stietz@stanford.edu

Anh Ta
UnifyID AI Labs
Redwood City, CA
ata@unify.id

ABSTRACT

In this paper, we'd like to disseminate a positive, serendipitous result entailing deep time-series transfer learning in the domain of human kinematics. We begin with a convolutional neural network model pre-trained to classify human subjects based on bipedal gait data emanating from tri-axial accelerometric sensors on commercial smartphones. We then use this model to transfer-learn into the problem domain of classifying subjects using tri-axial gyroscopic data emanating from a commercial motion-sensor device mapping to the activity of body-weight squats (of vividly different temporal spans). We achieve 87% top-1 accuracy on a 20-class problem even with a 60-40 train-test split. We hope that this cross-sensor, cross-device and cross-activity transfer learning success will pave the way for widespread deployment of deep transfer learning techniques in the domain of human kinematics analysis.

KEYWORDS

deep neural networks, gait, motion sensor time series

ACM Reference Format:
Vinay Prabhu, Stephanie Tietz, and Anh Ta. 2019. Classifying humans using deep time-series transfer learning : accelerometric gait-cycles to gyroscopic squats. In *MileTS '19: 5th KDD Workshop on Mining and Learning from Time Series, August 5th, 2019, Anchorage, Alaska, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Performing longitudinal studies of a cohort of athletes or patients becomes challenging, especially when they share the same medical equipment or sensors. Currently, it is being done manually via careful registration and tracking. In this paper, we address this challenge by performing machine-learning-aided automated classification of the cohort members by harnessing deep transfer learning. Researchers have enjoyed a lot of success recently in identifying humans at a large scale by performing deep learning on accelerometric gait data such as the data found in `GaitNet` (See [7, 8, 10]). Using these gait-trained model(s), we sought to discover if we can use transfer learning to bring human classification into the *squats domain.*

*Both authors contributed equally to this research.

MileTS '19, August 5th, 2019, Anchorage, Alaska, USA

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1.1 Background on Gait and Squat analyses

Studies that analyze gait and other human motion generally rely on video capture or inertial measurement unit (IMU) sensors [1, 7]. IMU sensors, like the one used in this study, have accelerometers and gyroscopes as well as additional sensors such as magnetometers, depending on the model. Smartphones have also been used as IMU sensors due to the fact that they have accelerometers and usually gyroscopes as well. In fact, the accelerometric gait data used to capture the data found in `GaitNet` was from smartphones [7]. IMU sensors have also been used previously with squat data in particular.
Physiotherapists and fitness coaches both work on squat mechanics with their athletes and patients, agreeing for the most part on what a "correct" squat should look like. They can also recognize common incorrect squats, or deviations, in their clients. Both professions often deal with groups, which makes real-time feedback more difficult, and athletes/clients also perform these exercises when alone. To this end, researchers used IMU sensors to help classify these movements [5]. After standardizing the correct squat form and some deviations, the researchers had different people do the same mechanics for each variation and achieved a multiclass model accuracy of 56.55% - finding the *squat variation* amongst the data. Prior to our work, to the best of our knowledge, no one has seen if a model could instead find the *person* amongst the data. We wanted to see if the model could locate common artifacts across all of a user's squat variations - artifacts that were unique to just one person. By achieving this goal, the model could assist physiotherapists and coaches by automatically identifying the user from a group of clients or athletes.

1.2 Transfer learning: The ImageNet to GaitNet analogy

As evinced in [5],[2],[3], the landscape of machine learning classification algorithms in the context of sensor-data driven human kinematics analysis is dominated by shallow learning algorithms entailing hand-crafted time and frequency domain feature engineering. One reason why deep learning techniques might not be as prevalent here is the lack of large volume of human kinematics sensor data to train the deep-net on. In the domain of computer-vision, this issue is circumvented by using transfer learning techniques that require small amount of domain-specific data to fine-tune a pre-trained model which was trained on a larger *general purpose* dataset such as ImageNet (See [9] for a survey of the field). In this paper, we disseminate the first results in the motion-sensor based himan kinematics time-series domain, where we showcase how we can use a pre-trained model- `DeepGaitID` - trained on `GaitNet` (analogous to ImageNet in computer vision) and transfer not just into other niche problem domain but also across different sensors (accelerometer to gyroscope). In addition to the high classification accuracy that we finally achieve and the low amount of training data required, there is one other facet of our results that was especially noteworthy. While the gait models were trained on real-world tri-axial accelerometric gait data emanating from the IMU sensors in commercial smartphones, the gyroscopic squat data was collected using an off-the-shelf IMU sensor device `metamotionr`[1]. Thus, the transfer learning is happening across different kinematic and sensor modalities at once.

The main contributions of the paper are as follows:

(1) Introduce a new tri-axial gyroscopic dataset (`Squat-20`) for the time-series community
(2) Disseminate the first ever research effort to identify humans on the basis of their squatting motion
(3) Successfully showcase deep transfer learning across two different modalities of human motion - bipedal gait and body-weight squats
(4) Open source the data and code associated with the experiments

In Section 2, we present the dataset collection procedure as well as detailed explanations into the type of data collected. Then in Section 3, we present the transfer learning procedure and results. Finally in Section 4, we conclude the paper and cover some of the directions in which we are currently extending this work.

2 `SQUAT-20`: DATASET DESCRIPTION

The data collected from participants was in the form of bodyweight squats, which are squats that are performed without any equipment (no extra weight). We asked the volunteers to do eight different variations of a squat: "correct" squats - those done with proper mechanics - as well as seven popular deviations (which we hope to use as a further classification environment in the future). The variation of squat movements performed can be seen in Table 1 along with the corresponding squat form ID used in the data.

Because there was not a constraint on squat experience, we noticed that some participants' data was more "noisy" than others. Visually, this noise was discernible to the trainer on-site as the more amateur participants, though accurately performing the squat with the instructed mechanics, often hesitated in their movements. These hesitation marks presented a challenge to classifying squat form using traditional batch classification algorithms as the unique squat form characteristics appeared to be harder to discern. We open up this dataset for outside exploration, and we will also continue our own exploration into joint *user-ID* and squat form classification. This noise, however, did not hinder participant-based classification.

2.1 Squat Form and Deviations

A correct squat is one that maximizes power efficiency through a person's global and local centers of gravity. As one deviates away from this ideal model, the movement becomes more inefficient and can often cause injury. Participants were instructed to follow the guidelines established by the National Strength and Conditioning Association (NSCA) [4]. With feet shoulder-width apart, the participant was told to sit back and let their knees slowly bend while keeping a flat back and their chest up. Their heels remained on the floor with the knees in line with their toes. When beginning their ascent, the participants made sure to extend their hips and knees at the same rate and to keep their chest up, again with their heels down and knees moving in the same line as their toes.

Squat Form Variations

SquatFormID	Description
0	Correct
1	Knees pass over toes
2	Knees move towards each other
3	Knees move away from each other
4	Heels up during movement
5	Hips shift to the left
6	Hips shift to the right
7	Improper hip flexion

Table 1: Every variation of the squat movement along with the corresponding form ID# used in the data

Some incorrect variations occur due to deviations from the body's global center of gravity. For example, a user may shift their hips to the left or right while descending into their squat (SquatFormID 5,6). Another deviation in this category occurs when a user shifts their mass too far forward, often seen through the user's heels raising off the ground (SquatFormID 4) or their knees passing far over their toes (SquatFormID 1). Improper hip flexion occurs when the user drops their chest during descent and/or leads the ascent by straightening their knees, inefficiently shifting their center of gravity during the movement (SquatFormID 7).

Other deviations displace local centers of gravity in the joints, specifically the knees. If the knees move towards each other or away from each other during the squat (SquatFormID 2,3), the user is away from their maximum potential which is when the knees remain tracked in line with the toes.

The set of deviations used for data collection is not comprehensive, but does reflect seven of the most popular deviations seen as verified by a physiotherapist and a personal trainer.

2.2 Data Collection

Twenty healthy participants agreed to contribute data for this study. Volunteers varied widely in age, from 22 to 60 years old with a median age of 27 and a standard deviation 9.35. They also varied in their history of squat performance from those with little to no experience up through advanced athletes.

After the participant signed a consent form, an Mbient Lab MetaMotionR IMU sensor was placed on their back at the L4 vertebra. Each participant was instructed on the expected form for the correct squat and each deviation. They performed ten correct squats and three to five of each deviation under the supervision of a certified personal trainer and fitness coach.

[1]https://mbientlab.com/metamotionr/

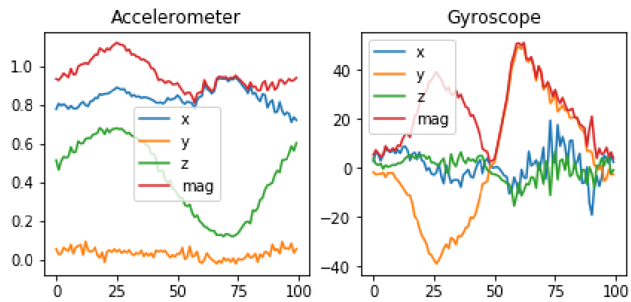

Figure 1: Visualization of a correct squat by volunteer-0

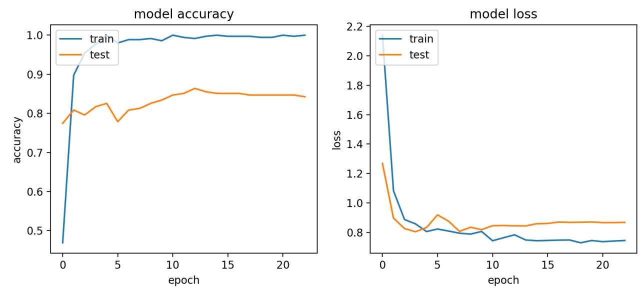

Figure 2: The epoch-wise plot of the train/test accuracy/loss

Accelerometric and gyroscopic data were collected at a frequency of 100Hz during each of the squats. An example of the tri-axial gyroscopic and accelerometric data tensors collected during a correct squat by volunteer-0 is shown in Fig 1.

Gyroscopic magnitudes for each participant over their squat variations can be seen in Figure 7. Data was then labeled according to squat form ID (from 0 to 7) as well as participant ID (from 0 to 19).

3 TRANSFER LEARNING PROCEDURE AND RESULTS

The `GaitNet` dataset is the largest accelerometric human gait dataset ever compiled [7, 10]. The entire dataset is a $1.2e6 \times 4 \times 100$ tensor and contains tri-axial accelerometric data collected from 1000 volunteers in 150+ countries. Each gait cycle matrix G is of size 4×100. The 4 axes are x, y, z plus the *magintude* axis. The dimension of the temporal axis is the end-result of resampling all gait cycles to size 100. That is, $G = [g_x(t), g_y(t), g_z(t), g_{mag}(t) = \sqrt{g_x^2(t) + g_y^2(t) + g_z^2(t)}]; t = 1, .., 100$. One of the models used to analyze this dataset was a deep CNN architecture named `DeepGaitID` that achieved 63% top-1 accuracy on the 1000-class GaitNet dataset. We chose this model to serve as our pretrained model for transfer learning.

As evinced in [9], there are many sub-frameworks within transfer learning. Generally in classification problems, the last layer of the original deep-net model is removed and replaced with a layer specific to the new data. Other layers can be frozen or left unfrozen. If all pretrained layers are frozen, the weights from those layers are used to extract features from the new data as that data is passed through to the added layer(s). High accuracy classification from this scenario is achieved if there is a high correlation between the original and new data. On the other end, all of the layers can remain unfrozen. In this case, the pretrained model acts as a weights-initializer for the new data [6]. For this study, we chose to retrain all layers of the original model. Some of the benefits of transfer learning, which we saw in this study, are that less data is required to achieve significant results and that the model learns faster than when it trained the original data [6].

In order to investigate the potential of transfer learning using this model, we removed the 1000-node softmax layer from the `DeepGaitID` model(froze every layer) and passed the `Squat-20` tensors to obtain 80-dimensional feature vectors. Figure 3 shows the t-SNE visualization of the features obtained colored according to the user-id. Clusters emerged, dividing users reasonably well, which sets the stage for transfer learning.

We constructed a `DeepSquatID` CNN by inheriting all the pre-softmax layers of the `DeepGaitID` model and introducing two new fully connected layers interspersed with dropout layers (with dropout rates of 0.25). The dimensionality of the 2 new layers were chosen to be 256 and 128 respectively. This was capped with a final softmax layer with 20 nodes and is as shown in Fig 4.

We then retrained this model on the `Squat-20` dataset using categorical cross-entropy loss and the `rmsprop` optimizer. To provide regularization, we used early stopping and reduced learning rate on plateau (factor=0.1, $\epsilon = 1e-4$) strategies. We also performed label smoothing as an additional regularization pre-processing step with a smoothing factor of $\epsilon = 0.1$. It is worth noting that we made no effort to normalize the data in order to bring the gyroscopic data (whose tri-axial readings, that is the roll, pitch and the yaw, are in rad/sec) to the accelerometric world (whose measurements are in ms^{-2}).

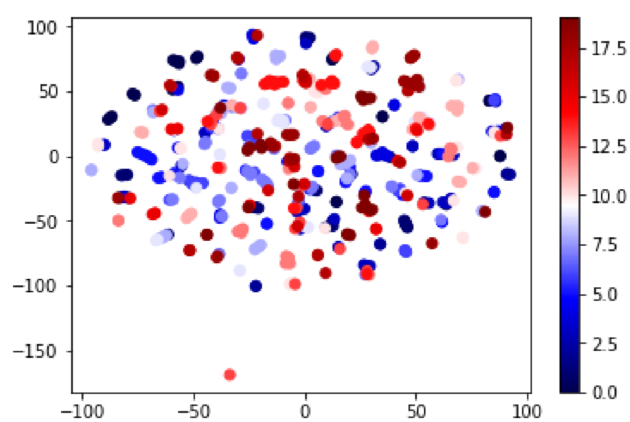

Figure 3: tSNE visualization of the 80-dimensional features extracted from the pre-softmax layers of the `DeepGaitID` classifier. The colors denote the *user-ID*

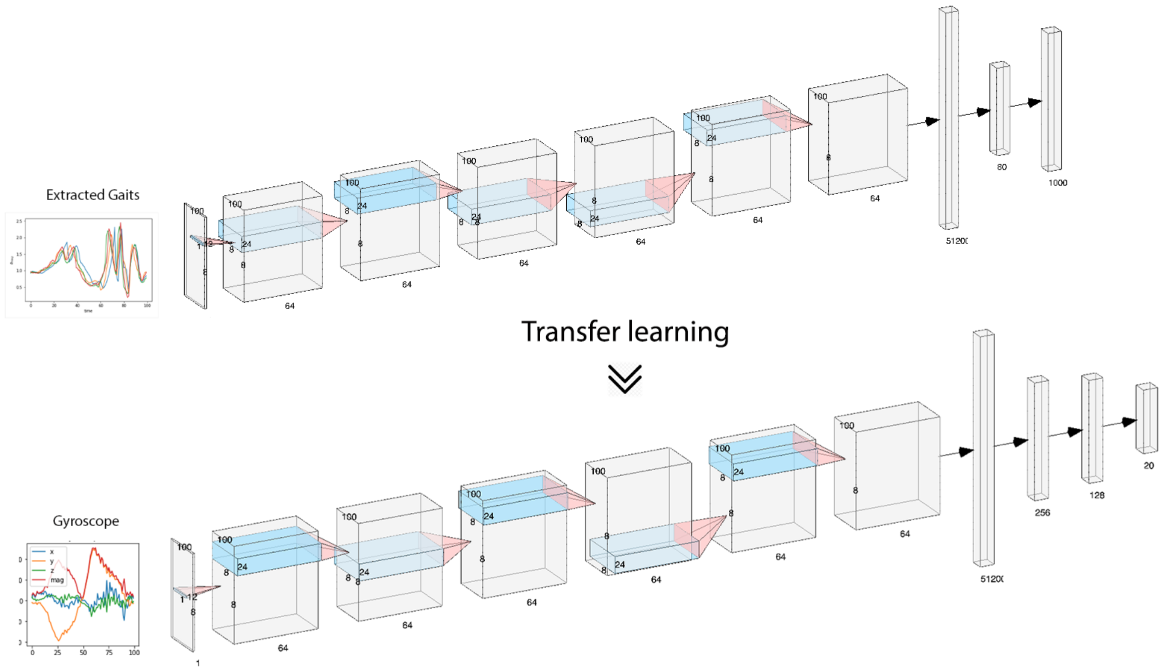

Figure 4: The post-transfer learning `DeepGaitID` architecture

In order to ensure fast enrollment of the athletes/patients, we performed a rather frugal train-test split of 0.6 : 0.4 which left us with 352 tensors in our training set and 235 tensors in our testing set. Even with this small amount of data, our model achieved a 87% top-1 accuracy as seen the classification report in Table 2. In Figure 5, the class-wise confusion matrix reveals a low number of misclassifications and high accuracy across the classes. For further exploration into this study, a *colab* notebook[2] showcasing the obtained results has been duly open-sourced.

4 CONCLUSION AND FUTURE WORK

In this paper, we were able to showcase a successful transfer learning experiment that entailed using a deep CNN model pretrained on the state-of-the-art `GaitNet`. Though the original dataset contained accelerometric gait data collected from commercial phones, we were able to transfer human classification into the domain of gyroscopic squat exercise signatures emanating from a commercial off-the-shelf IMU sensor kit. This is currently a work in progress and we are extending this work in the following two directions:

(1) Trying to replicate the results with tri-axial magnetometric data.
(2) Performing participant classification in conjunction with squat-type classification. A naive attempt at using the `DeepGaitID` CNN and trying to predict the squat-type rather than user-ID yielded an accuracy of $\sim 45\%$.

[2]https://bit.ly/2VG6dXP

REFERENCES

[1] A. Ahmadi, E. Mitchell, F. Destelle, M. Gowing, N. OConnor, C. Richter, and K. Moran. 2014. Automatic Activity Classification and Movement Assessment During a Sports Training Session Using Wearable Inertial Sensors. *2014 11th International Conference on Wearable and Implantable Body Sensor Networks* (2014).

[2] Kun-Hui Chen, Po-Chao Chen, Kai-Chun Liu, and Chia-Tai Chan. 2015. Wearable sensor-based rehabilitation exercise assessment for knee osteoarthritis. *Sensors* 15, 2 (2015), 4193–4211.

[3] Zhenghua Chen, Qingchang Zhu, Yeng Chai Soh, and Le Zhang. 2017. Robust human activity recognition using smartphone sensors via ct-pca and online svm. *IEEE Transactions on Industrial Informatics* 13, 6 (2017), 3070–3080.

[4] J. Coburn and M. Malek (Eds.). 2012. *NSCA's Essentials of Personal Training*. National Strength and Conditioning Association.

[5] M. O'Reilly, D. Whelan, C. Chanialidis, N. Friel, E. Delahunt, T. Ward, and B. Caulfield. 2015. Evaluating Squat Performance with a Single Inertial Measurement Unit. *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (2015).

[6] S.J. Pan and Q. Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010), 1345–1359.

[7] Vinay Uday Prabhu and John Whaley. 2009. GaitNet 1.0 : State-of-the-art accelerometric gait based identification of humans. In *Big Data in Precision Health Conference*. Stanford.

[8] Sebastijan Sprager and Matjaz Juric. 2015. Inertial sensor-based gait recognition: A review. *Sensors* 15, 9 (2015), 22089–22127.

[9] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. 2018. A Survey on Deep Transfer Learning. *arXiv preprint arXiv:1808.01974* (2018).

[10] Daniel Wen Vinay Uday Prabhu and John Whaley. 2018. SIMUni: Sampling Impostors from Misfit Universal Background Models in accelerometric gait biometric verification. In *BayLearn: Bay Area Machine Learning Symposium*.

A APPENDIX

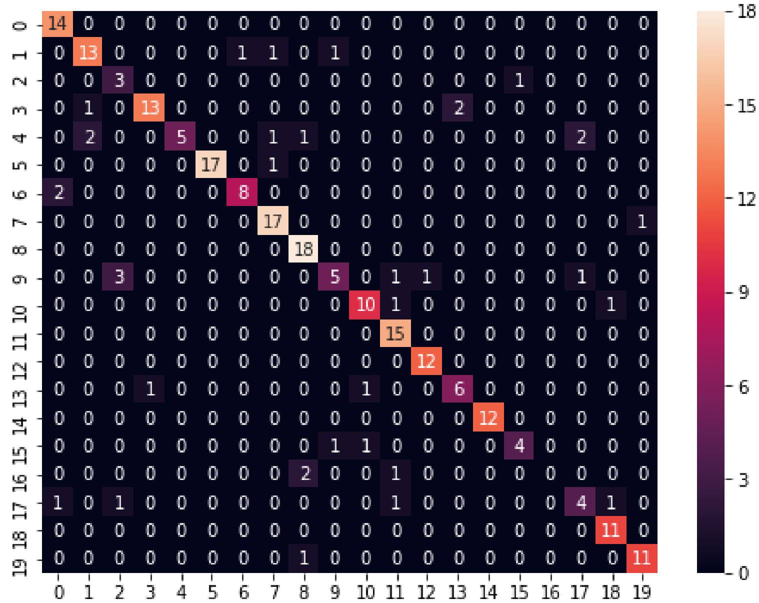

Figure 5: Confusion matrix for the `Squat-20` dataset

	precision	recall	f1-score	support
0	0.82	1.00	0.90	14
1	0.81	0.81	0.81	16
2	0.43	0.75	0.55	4
3	0.93	0.81	0.87	16
4	1.00	0.45	0.62	11
5	1.00	0.94	0.97	18
6	0.89	0.80	0.84	10
7	0.85	0.94	0.89	18
8	0.82	1.00	0.90	18
9	0.71	0.45	0.56	11
10	0.83	0.83	0.83	12
11	0.79	1.00	0.88	15
12	0.92	1.00	0.96	12
13	0.75	0.75	0.75	8
14	1.00	1.00	1.00	12
15	0.80	0.67	0.73	6
16	0.00	0.00	0.00	3
17	0.57	0.50	0.53	8
18	0.85	1.00	0.92	11
19	0.92	0.92	0.92	12
micro avg	0.84	0.84	0.84	235
macro avg	0.78	0.78	0.77	235
weighted avg	0.84	0.84	0.83	235

Figure 6: The classification report of the Squat-ID classification problem

Figure 7: Visualization of the variations in g_{mag} across different users performing squats [User-(userID) (squatformID)]