PREDICTION OF GAIT PARAMETERS FROM SMPL MODEL

Lucas Mørch Emcken (s214625), Mette Breuning Nielsen (s174152) and Moamal Fadhil Abdul-Mahdi (s195398)

DTU Compute, Technical University of Denmark

ABSTRACT

We present a learned model that is able to estimate the gait parameters double support time, stride length and asymmetry from videos. Our method takes input parameters for the rotation angle of joints (X, Y and Z) from the Skinned Multi-Person Linear (SMPL)[1] model and use them to train a model on foot to ground contact. Most existing methods need an on body device or the need of professional guidance to collect data and gait parameters where our method circumvents this by calculating these parameters only from a video of an individual walking from the side. The model achieves a contact estimation accuracy of 93%, and a mean absolute error of 2.24% for double support, 13.3cm for stride length, and 1.7% for assymetry.

Our code, model and demo is available on GitHub: `https://github.com/TheLucanus/DLGaitAssessment`

Index Terms— SMPL, 4D humans, gait analysis, gait parameters, video analysis

1. INTRODUCTION

Assessing simple gait parameters can help subjects gain valuable knowledge about their health[2]. However, it can be rather difficult to get a fairly accurate assessment of these parameters without professional help, time consuming video analysis or smart devices such as a smartwatch or smartphone [3]. Our approach uses 4D humans[4] to extract a SMPL model that tracks the rotations of joints through time, which coupled with a neural network allows us to estimate foot contact with the ground. This foot contact along with positional data from 4D humans allows us to automatically analyze the double support time, stride length and asymmetry of a gait. This method provides easier gait analysis and could potentially lead to early detection of various health issues, improved rehabilitation strategies, and overall better quality of life.

2. RELATED WORK

Walking parameters with on body device. Countless studies have been done to assess walking parameters. Such works have often focused on retrieving these parameters through an on body device such as health apps on your phones. This makes the data easy accessible for each individual. We retrieve these parameters without the need of an on body device thereby making it possible to analyse peoples gait parameters all around the world with just one video.

Gait parameters through video analysis of a TUG test. Extracting gait parameters through video analysis has been done before to retrieve knowledge on cognitive-motor interference in dual-task testing.[5] The study did video analysis on time up and go (TUG) tests to extract gait parameters. They ended up with a reliable method to retrieve gait parameter and asses the individuals health. Their course of action is very similar to what we do however our study access the gait parameters though a simple walk/run thereby eliminating the need for a TUG test and the need for professional assistance.

3. MODEL AND MATERIALS

To assess gait parameters such as double support time and stride length, it is sufficient to detect when a foot is touching the ground along with their translation in space. This is because the gait parameters can be derived from this data, essentially creating a binary classification problem for foot contact in computer vision. Therefore, we trained a deep learning model on videos of individuals walking to detect the exact frames each foot is touching the ground.

Instead of convolving over every frame in a video sequence, we instead gather information about each joint during gait. To do this, we used 4D humans. 4D humans is a fully transformer based approach for recovering 3D meshes of human bodies. To create human meshes, 4D humans use SMPL which along with a mesh provides a rotation matrix for 23 different joints in the body. The SMPL model from 4D humans not only offers rotation matrices relative to each other, but also positional coordinates, making a subject's movement in space important. This means that gait on a treadmill cannot be analyzed accurately, since the subject is essentially standing still. Therefore, to get the best results and accurate tracking, the entire body must be within the frames or at least the body parts associated with a joint, and the subject must walk through space.

To predict when the feet were touching the ground, we used a fully connected feed forward neural network (FFNN). Our FFNN consisted of 69 input parameters, (X, Y and Z rotation angle for the 23 joints from the SMPL model). The FFNN had 1 hidden layer of 64 nodes and the output layer consisted of two nodes for binary classification. We used the ReLU activation function between the layers and a sigmoid activation function in the output layer as sigmoid activation function is suitable for binary classification.

To train our FFNN, we had to gather and annotate data. So we created a data set of 24 recordings of people walking from the side with a static camera. We manually annotated what frames the corresponding feet were in contact with the ground.

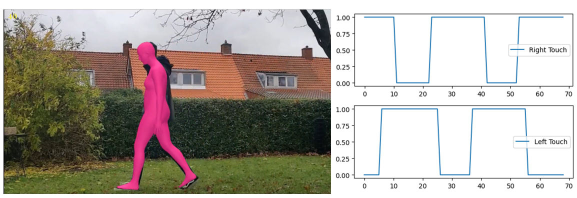

Fig. 1. Ground truth for foot contact (Y-axis) over frames(X-axis). 1 Corresponds to contact, 0 corresponds to no contact.

Furthermore, we used Logger Pro on the videos to generate a ground truth for the stride length and asymmetry. We split our data into a training set with 19 of the recordings and a test set with the remaining 5 videos.

We used binary cross entropy loss as our loss function since our problem is of binary classification nature. Binary cross entropy is appropriate since we are classifying weather or not each foot (left or right) is in contact with the ground in a single video. The equation for binary cross entropy can be seen in equation 1 in appendix.

3.1. Training the FFNN

We chose the Adam optimizer for our simple FFNN due to its adaptive learning rate, which helps prevent overfitting by assigning individual learning rates to each weight. Additionally, we employed dropout and batch normalization as extra measures against overfitting, although they had minimal impact on the FFNN's performance.

We trained our FFNN for 30 epochs to make sure the it was trained for an adequate number of times. During these epochs the loss steadily decreased until around the 20th epoch where the test loss started flattening. This can be seen in Figure 12 in appendix.

3.2. Data augmentation

The only method we used for getting more data was to record more videos, as altering the speed of videos would not generate new data. This is because the stride length is also determined by SMPL coordinates pr. frame. This data argumentation method does not change the number of frames and therefore it will not have an effect on the result. It might also be possible to mirror the SMPL model to essentially get 2 training sets from 1 video, however we did not employ this tactic.

3.3. Estimating double support time, stride length and asymmetry

- **Double support** can be calculated as the percentage over a segment where both feet contacting the ground.
- **Stride length** can be estimated by using the right or left foots contact, we can construct a phase where the distance the subject travels during that phase is corresponding to the stride length of that leg.
- **Asymmetry** can lastly be calculated as the percentwise difference between left and right stride length

4. RESULTS

The FFNN is used to predict when each foot is in contact with the ground. Figure 2 shows the model prediction (blue) and the ground truth found by manual detection (orange) for each foot (right and left). As can be seen from the results shown on Figure 2, the model performs very well, with an overall accuracy at 93% across the full test set compared to a baseline accuracy (estimating both feet are in contact all the time) at 69%.

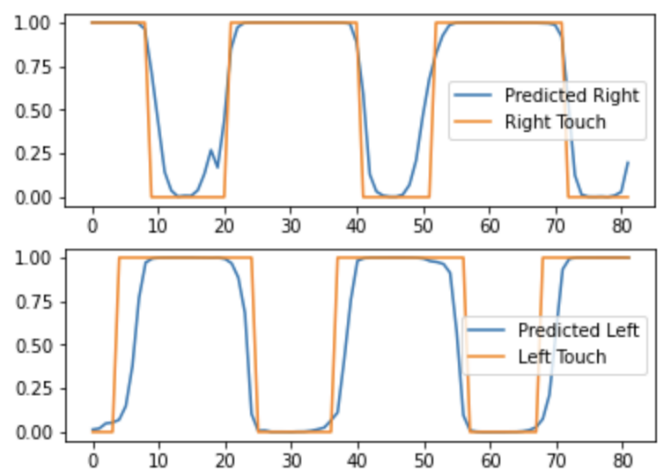

Fig. 2. Prediction of foot to ground contact found with the FFNN model (blue) and the manually detection (orange) for right and left foot. Y-axis: 1 indicates contact- and 0 indicates no contact with the ground. X-axis: frames in the video.

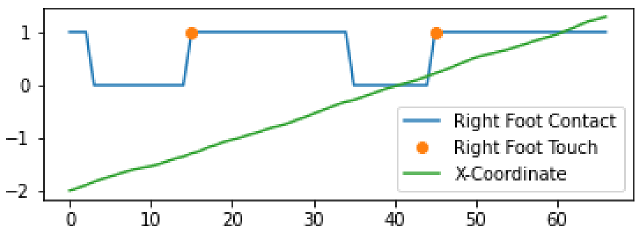

Fig. 3. Estimation of stride length by X-coordinate over a contact phase, Y-axis: Contact and X-translation, X-axis: Frames

The stride length can then be estimated from the phase of contact (as marked by the orange dots on Figure 3), where the change in X corresponds to the stride length. We noticed a scale difference between the 4D humans estimated X distance and the measured one, corresponding to a linear relationship, so the X distance is scaled up accordingly.

On Table 1, we can see the performance across the various gait metrics compared to a baseline, that estimates all of the parameters as the mean values of the training set.

	MAE	STD	MAE Baseline	STD Baseline
DS	2.24 %	1.34%	3%	1.98%
SL	13.3 cm	11.7 cm	21.6 cm	12.8 cm
Asy	1.7%	1.05%	0.526 %	0.428%

Table 1. Model performance (MAE = Mean absolute error, STD = Standard deviation) across gait metrics (DS = Double support, SL = Stride length, Asy = Asymmetry)

From the data we can see that our method of estimating gait parameters performs better than the baseline in assessing double support time and stride length, however, the asymmetry measurement is off by quite a bit. This is because asymmetry is measured from the difference in stride lengths, and is typically between 0-2%. As such even a small error in stride length or contact can propagate, and lead to a large error in asymmetry. For example if a contact point is estimated a frame after the true contact, the person might have moved several centimeters, leading to a large asymmetry.

5. DISCUSSION

5.1. Leg switching

During the training of our model, we discovered that the SMPL model would sometimes switch between the legs during the tracking from one frame to another as illustrated on Figure 4.

Fig. 4. Leg switch between one frame and the next

To mitigate this problem, we first extracted the rotations of the joints in the legs of the SMPL model (hip, knee, ankle and foot) and calculated the euclidean distance between these rotations for the current frame and the next. The idea here is that a large difference in rotations between two frames would correspond to that of two legs switching place, this is seen in Figure 5.

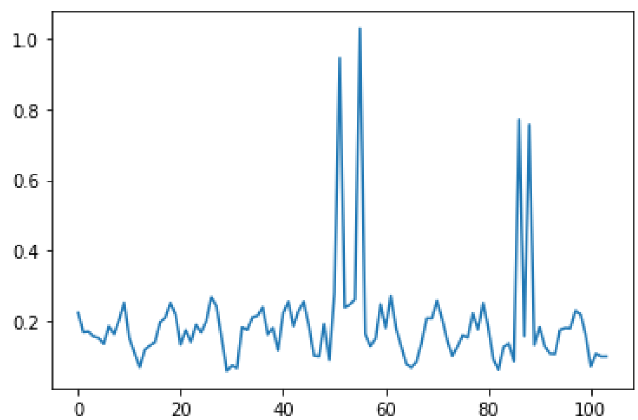

Fig. 5. Euclidean distance between the rotation in the leg joints compared to next frame (Y-axis), X-axis: Frames

In this plot we see two occurrences of leg switches, corresponding to pairs of peaks, where the first peak would be the leg switching initially, and the second being the leg switching back into place. With this graph, we extracted the peaks at values larger than $\mu + 2\sigma$ of the initial data. Then re-interpolated these points linearly from the data at endpoints of the interval, resulting in the graph shown on Figure 6.

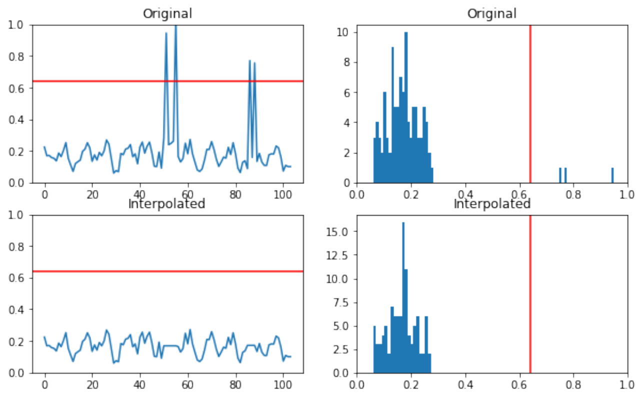

Fig. 6. Re-interpolation of points with leg errors

Using the model to predict the leg positions, we notice a significant improvement in performance, as can be seen by the new contact point estimations on Figure 7.

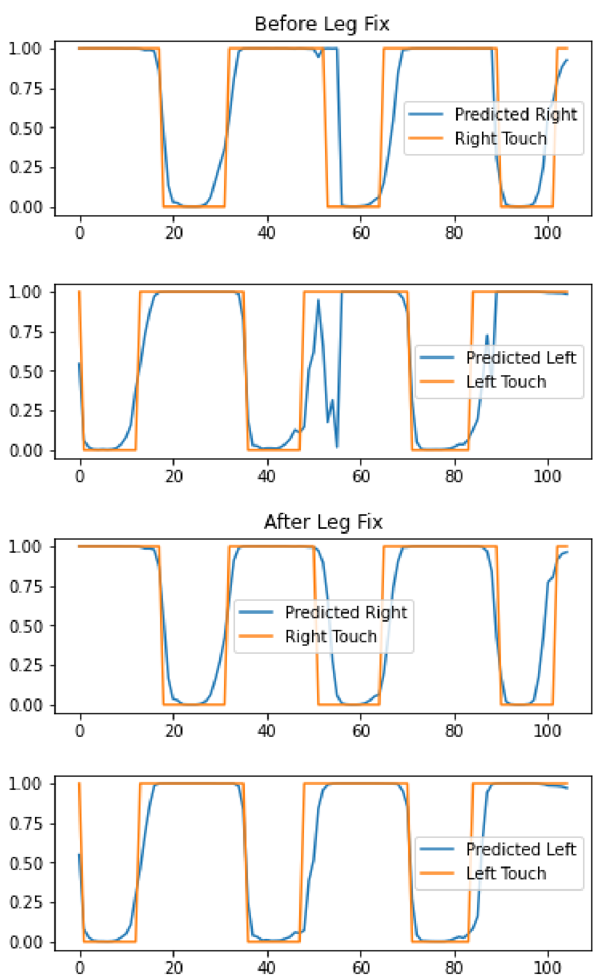

Fig. 7. Contact prediction on legs before and after fix

While this method catches many cases of leg switching, there are still instances where it does not work. There is also a limit on the maximum interval between peaks, as interpolation over many frames would be too destructive.

5.2. Stress testing

As our model is only trained on side shots of people walking on a flat ground (where one foot is always touching the ground), the model has some trouble generalizing to other use cases such as stairs or running, as such we have conducted a small stress test to check the robustness of the model.

Stairs

When walking on stairs, the contact points to the ground would be skewed, since you (from a pose perspective) touch the ground earlier when walking up stairs, and later when walking down stairs.

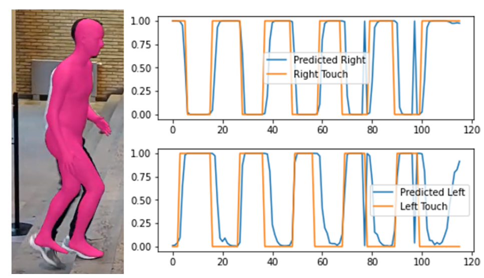

Fig. 8. Walking up stairs, Model contact accuracy: 82%, Baseline contact accuracy: 48%

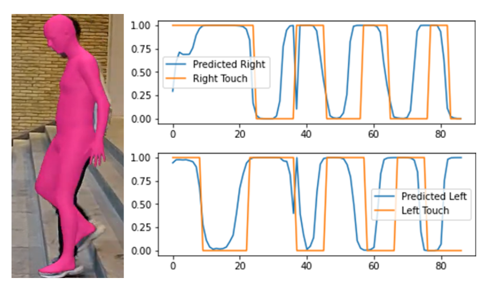

Fig. 9. Walking down stairs, Model accuracy: 75%, Baseline accuracy 53%

Interestingly when walking down the stairs, the models prediction of contact points seem to be leading the actual contact points a bit. Since the model is trained on flat ground data, the point where the model expects the feet to touch the steps are above the actual steps, making the model predict the contact points before they actually happen. A similar opposite effect appears on graph for walking up the stairs, however since the sequence is longer, the effects are not so noticeable. Both double support times are rather imprecise (Up: predicted 13%, actual 3.4%. Down: predicted 26.4%, actual 13.7%)

As visible on both graphs on Figure 8 and 9, the SMPL model encountered leg switching, where the leg fixing algorithm could not detect and fix the cases effectively.

Running

When a persons is running, there is often instances where no foot is touching the ground, ("flight phase" or "non-support phase"). Such a phase is not present in the training set as it only contains videos of people walking, where there is constant support from at least one foot.

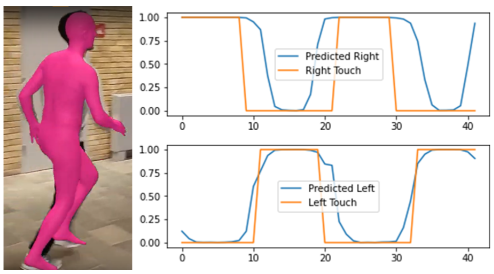

Fig. 10. Running, Model accuracy: 83%, Baseline accuracy 42%

When running, the model seems to still be able to capture the contact points with only a slight reduction in accuracy as shown on Figure 10. However the double support time is completely off. During running, it is uncommon to have any double support, and instead have a flight phase, however the model still predicted a double support time of 14.3%, despite the actual double support being 0%

Jumping jacks

While performing jumping jacks, the SMPL parameters would be much different than what the model has been trained on, as there is next-to-no point during a normal walk cycle where a persons pose mimics that of a jumping jack.

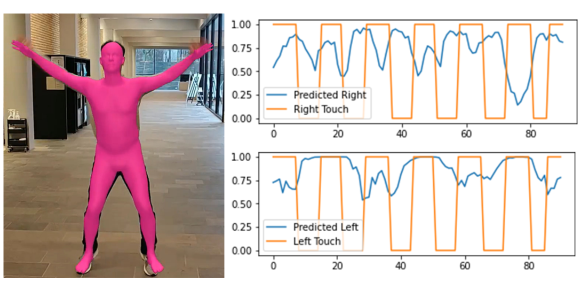

Fig. 11. Jumping jacks, Model accuracy: 53%, Baseline accuracy 56%

As a result, the model cannot even outperform the baseline accuracy in estimating contact points as shown on Figure 11.

5.3. Future work

For the future we would like to apply Simultaneous Localization and Human Mesh Recovery (SLAHMR) [6] instead of 4D humans, since this would let the model be able to capture features with a dynamic camera, enabling our model to work in the wild. Furthermore, it is possible that 4D humans has difficulties estimating whether its a larger person further away or a smaller person closer, which could lead to deviations in the change of position. Introducing, SLAMHR, to the model would be a way to work around this. However, due to time constraints we were unable to implement this and it is therefore one of the major parts we would focus on in the future. Furthermore it would be ideal to collect more data for the model to train on. We would also like to improve the leg fixing algorithm to capture more cases of legs switching place, which could further improve the model.

6. CONCLUSION

Using 4D humans to extract SMPL parameters from a video, we have applied a FFNN model to estimate right and left foot contact, double support, stride length and stride asymmetry. For estimating touches, we received an accuracy of 93%. Using this to estimate the gait parameters, we got a mean absolute error of 2.24% for double support, 13.3 cm for stride length and 1.7% for stride Asymmetry. Estimating Asymmetry was more error prone due to it being affected a lot by small changes in stride length, and we foresee that using SLAHMR instead of 4D humans will improve tracking quality, reducing the magnitude of the problems, as well as allowing for a dynamic camera to better capture more space requiring movements like running. Furthermore, we saw after running stress tests on the neural network that the model performs best on videos where the individuals are walking on a flat surface which is unsurprising, since it is the same types of data the model is trained on. Applying the model to untrained cases such as running and using stairs, we noticed a drop in accuracy (83% for running, 82% for walking up stairs, 75% for walking down), as the model is not trained on these actions. Lastly when performing jumping jacks, the model performed significantly worse at a 53% accuracy compared to 56% baseline.

7. REFERENCES

[1] J. Romero G. Pons-Moll M. J. Black M. Loper, N. Mahmood, "Smpl: A skinned multi-person linear model," *ACM Transactions on Graphics*, vol. 34, no. 6, 2015.

[2] A. Zech A. Stotz, D. Hamacher, "Relationship between muscle strength and gait parameters in healthy older women and men," vol. November, 2022.

[3] Apple, "Measuring walking quality through iphone mobility metrics," vol. May, 2022.

[4] R. Jathushan K. Angjoo-M. Jitendra G. Shubham, P. Georgios, "Humans in 4d: Reconstructing and tracking humans with transformers," 2023.

[5] H. B. Åhman-O. Tarassova A. Arndt V. Giedraitis L. Berglund K. Halvorsen A. C. Åberg, F. Olsson, "Extraction of gait parameters from marker-free video recordings of timed up-and-go tests: Validity, inter- and intra-rater reliability," *Gait and Posture*, vol. 90, pp. 489–495, 2021.

[6] J. Malik A. Kanazawa V. Ye, G. Pavlakos, "Decoupling human and camera motion from videos in the wild," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2023-June, pp. 21222–21232, 2023.

8. APPENDIX

1.

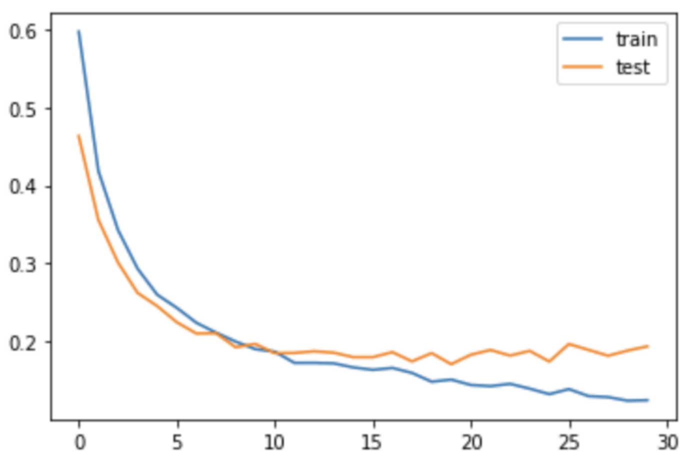

Fig. 12. Loss curve for training and testing

2.

$$\ell(x, y) = L = \{l_1, \ldots, l_N\}^\top \tag{1}$$
$$l_n = -w_n \left[y_n \cdot \log x_n + (1 - y_n) \cdot \log(1 - x_n) \right]$$

$$\ell(x, y) = \begin{cases} \mathrm{mean}(L), & \text{if reduction} = \text{'mean';} \\ \mathrm{sum}(L), & \text{if reduction} = \text{'sum'.} \end{cases} \tag{2}$$