# GRF-MV: Ground Reaction Force Estimation from Monocular Video

Juni Katsu
jxk010@alumni.bham.ac.uk
Esha Daspupta
esha.dasgupta@gmail.com
Hyung Jin Chang
h.j.chang@bham.ac.uk

University of Birmingham
Birmingham, UK

**Abstract**

Estimating ground reaction forces from monocular video has essential applications in healthcare, such as rehabilitation, injury prevention, patient monitoring, and physical therapy. However, it is a challenging problem due to the complexity of human motion, the limited availability of training data, and the difficulty of estimating contact and forces from only 2D monocular video. This paper presents a novel approach for estimating ground reaction forces (GRFs) from monocular video by combining deep learning-based 3D human mesh recovery with physics-based optimizations. Existing techniques for measuring GRFs rely on specialized sensors or multiple camera setups, limiting their applicability outside of lab settings. In contrast, our proposed approach requires only a single video camera, making it suitable for deployment in sports, clinical and home environments. A deep neural network is trained to recover 3D human mesh parameters from each frame, which are further refined using physics-based optimization (HybrIK-XL). GRFs are then estimated from the 3D foot velocities and contact modeling. The approach is evaluated on the GroundLink dataset, demonstrating improved accuracy over prior methods. [1]

## 1 Introduction

Measuring and analyzing human motion and contact forces offers valuable insights for a wide range of healthcare applications, including rehabilitation, injury prevention, patient monitoring, and physical therapy. The ability to precisely measure forces exerted on the body is a vital component of this process, particularly Ground Reaction Forces (GRF), the force exerted by the ground on a body in contact with it. Although current approach rely on force plates or wearable pressure sensors, these approaches are typically expensive, restrict the range and fluidity of motions that can be captured, and are invasive to the patient[1].

Recent advances in Physics-based optimization have allowed the estimated forces and body kinematics of the data to be constrained to only physically plausible movements, offering new opportunities for non-invasive assessment[24][27][26]. Additionally, recovering

[1] The code is ready to be released.https://github.com/juniKRP/GRF_MV

the 3D body mesh from 2D video frames taken with a single (monocular) or multiple cameras provides a promising alternative for capturing human movement data. Parametric body models like SMPL[15] and SMPL-X[18] represent the 3D shape parameter and pose of the human body. However, recovering the 3D body shape, pose, and contact forces from 2D projected images is extremely challenging due to depth ambiguity, occlusions, variations in body shape and clothing, and the complexity of human motion.

This paper tackles the challenge of estimating GRFs from monocular RGB video for the purpose of the enhancement of healthcare by leveraging recent advances in 3D human mesh recovery and physics-based optimization with inverse kinematics approaches. We introduce a novel framework leveraging SOTA models to perform video-based GRF estimation using only monocular video.

Our approach first fits a parametric 3D body model (SMPL-X) to each frame of a monocular video using a novel deep neural network combination (HybrIK-XL). An initial estimate of the body's pose and shape is refined through a physics-based optimization methodology, ensuring realistic interactions with the ground plane. Finally, the GRFs are calculated from the optimized foot motions and evaluated using the GroundLink[7] dataset. We focus on monocular video of a single human subject performing dynamic motions.

Our method combines current state-of-the-art methods on 3D mesh recovery and estimating GRFs with improvements in performance, requiring less computational resources compared to existing methods. To demonstrate the estimated GRF, we use Unity[6] with character animations modified to use the generated keypoints for each frame of a single video. Evaluation of our model takes the accuracy of our calculated GRF on the captured motions from the GroundLink dataset.

## 2 Related Work

Estimating human motion and contact forces from videos has been a longstanding challenge within computer vision and graphics. We review the relevant background material and prior work on 3D human mesh recovery, physics-based optimization, and data-driven GRF prediction.

**3D Mesh Recovery:** While recovering the full 3D human shape and pose from a single image is a challenging problem, many previous works have contributed to the improvement of its accuracy using different deep-learning neural networks for whole-body pose estimation from videos. Parametric 3D body models like SMPL[15], and SMPL-X[18] are a compact parametric representation that can be predicted using deep learning neural networks. Methods like HMR[8] and SPIN[10] use CNNs to return body model parameters from the pixels of images directly. Recent approaches estimate 2D-keypoints of the human body in the frame and silhouettes as intermediate representations to improve accuracy. Combining 2D keypoint estimation with information inferred from recurrent networks (VIBE)[9], motion discriminator (TCMR)[3], or transformers (MeshTransformer)[14] as examples, shows significant improvement in the accuracy of the captured motion. However, most of these methods are trained on datasets of posed subjects and do not handle foot contacts or the depth of the camera in the environment.

Another work of interest is HybrIK[12], which fits the SMPL body model to each 2D-keypoint estimated frame of video using the HR network [21] and inverse kinematics (IK) optimization to enhance joint angles and twists, which potentially reduces foot sliding. In this paper, we propose an extended framework of HybrIK, HybrIK-XL, using HybrIK-X[13]

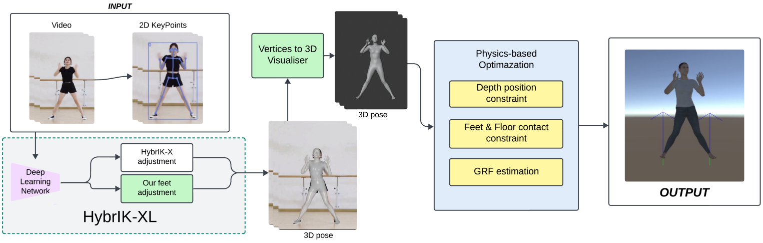


Figure 1: Pipeline overview of GRF-MV. Inputs can be any monocular video and its extracted 2D keypoints, then passed to the neural network to train and feet contact points are further adjusted based on the loss value (HybrIK-XL). Generated 3D poses are converted to an animation and optimized within Unity and calculated GRFs are visualised.

to fit and optimise the SMPL-X body model meshes as input to our GRF prediction pipeline.

**Ground Reaction Force Estimation:** The ground reaction force is the force exerted by the ground on a body. GRF provides valuable insights into human motion, balance, and the distribution of force along the body, making it essential for various healthcare applications. However, accurately measuring GRF remains a challenging task. Traditionally, GRF is measured using specialized equipment like force plates[5] or pressure-sensitive insoles[20][2].

Currently, several datasets have been developed to enhance the accuracy of GRF estimation and new datasets of motion capture with GRFs [11] and [28] datasets continue to emerge. However, these datasets are often limited to specific aspects of movement, target stages and diseases in subjects. Datasets that were considered for this approach were UnderPressure[17] and PSU-TMM100[25]. However, due to the fact that the UnderPressue dataset lacks 3D body shape information and the PSU-TMM100 dataset is incompotible with SMPL-X, they were deemed unusable.

The most suitable dataset for our approach is the GroundLink[7] dataset, which provides synchronized video, motion capture, and force plate data. The dataset includes 60 subjects performing a wide range of activities, such as stretching, walking, and running. A motion capture system was used to track the 3D body pose and shape, visualized using the SMPL-X model. The dataset also includes ground truth force plate data, making it suitable for evaluating our approach.

**Physics-Based Character Animation:** Accurately estimating foot contact and GRFs from monocular video is challenging due to 2D image ambiguity and complex human motion dynamics. [19] proposed a method that combines a deep neural network for 3D human motion estimation with a physics-based optimization module for foot contact optimization. Similarly, [24] introduced a method that estimates human motion by optimizing joint torques and contact forces to minimize discrepancies between observed and simulated motion.

Recent works have further advanced these techniques by integrating physics-based optimization with deep learning and sensor data. [26] proposed Physical Inertial Poser (PIP), a real-time motion tracking system combining sparse inertial sensor data with physics-based optimization. [27] introduced PhysDiff, a physics-guided human motion diffusion model that generates physically plausible motions by incorporating physics-based optimization into the diffusion process. [23] leveraged intuitive physics to improve 3D human pose estimation accuracy and plausibility. These approaches demonstrate the potential of combining data-driven methods with physics-based optimization to enhance the accuracy and realism of human motion estimation and synthesis from video or sensor data.

# 3 Methodology

This section describes our framework for estimating GRFs from a monocular video, summarized in Figure 1. Given an input video, human poses are estimated for each frame in 3D meshes using HybrIk-X [13] and refined within our HybrIK-XL framework, enhancing kinematic plausibility. An animation is then generated using the joint sequence and is then displayed within Unity. Finally, we apply the GRF equation [22] to the animated model to calculate and evaluate GRFs.

## 3.1 3D Mesh Recovery

For our framework, we have chosen HybrIK-X, a state-of-the-art model for 3D mesh recovery. Additionally, HybrIK-X is trained on the SMPL-X mesh dataset that allows the use of shape, expression, and pose parameters, enabling the recovery of a more detailed 3D human mesh.

HybrIK-X demonstrates impressive performance in estimating joint rotations and positions. However, significant errors in foot positions can still be observed. Accurate contact estimation between the feet and the floor is crucial to estimate GRFs. Therefore, we introduce HybrIK-XL, an additional loss term framework specifically designed to improve the accuracy of foot-ground contact estimation.

In order to obtain an accurate ground plane, we manually select a frame from the video where the person is standing on the ground, which is used to calculate the average of the z-axis coordinates of the right and left ankle in that frame and set it as the ground truth $G_z$. This value serves as a reference for comparison with the current feet position $F_z$ along the z-axis. The loss function is designed to adapt based on the relationship between $F_z$ and $G_z$.

$$\mathcal{L}_{feet} = \begin{cases} E_0, & \text{if } F_z - G_z - \text{offset} < 0. \\ E_1, & \text{otherwise.} \end{cases} \tag{1}$$

The loss function $E_0$ is designed to minimize the occurrence of feet clipping as much as possible. Equation 2 calculates the mean squared error (MSE) between the predicted feet positions $\alpha_k$ and the ground-truth feet positions $\hat{\alpha}_k$ for $K$ feet joints. This will encourage the predicted foot positions to align closely with the ground truth, effectively pushing the feet upwards to avoid clipping.

$$E_0 = \frac{1}{K} \sum_{k=1}^{K} \parallel \alpha_k - \hat{\alpha}_k \parallel^2 \tag{2}$$

In reality, the feet are not always in contact with the ground during various movements such as walking, running, or jumping. Thus, we cannot rely on a loss term that consistently pushes the feet towards the ground plane. Therefore, we introduce the loss term $E_1$, which is applied when the feet are predicted to be above or on the ground plane ($F_z - G_z - \text{offset} \geq 0$).

$$E_1 = -\sum_{k=1}^{K} \omega_0 log(1 - \bar{\alpha}_k) \tag{3}$$

where $\bar{\alpha}_k = (\alpha_k - \hat{\alpha}_k)/\sigma_k$ for $K$ feet joints and $\sigma_k$ denotes the standard deviation. Equation $E_1$ employs a log-likelihood loss to create a gentle loss curve that allows for the possibility of feet floating, controlled by the weight $\omega_0$. Using a log-likelihood loss, we ensure that the penalty for feet floating gradually increases as the distance between the feet and the ground plane grows.

**From pretrained model:** The HybrIK-X pretrained model contains three key components: pose, camera, and twist angle estimation. Additional parameters from HybrIK-X are:

$$\mathcal{L}_{pose} = -\sum_{k=1}^{K} logQ(\bar{p}_k) - logG_{\psi} + 3log\sigma_k \tag{4}$$

$$\mathcal{L}_{cam} = \parallel s^0 - \hat{s} \parallel^2 \tag{5}$$

where $\hat{S}$ is the ground-truth scale factor.

$$\mathcal{L}_{tw} = \frac{1}{K}\sum_{k=1}^{K} \parallel (cos\phi_k, sin\phi_k) - (cos\hat{\phi}_k, sin\hat{\phi}_k) \parallel^2 \tag{6}$$

**From SMPL-X model:** The SMPL-X additional parameters: shape $\beta$, expression $\psi$ and rotation $\rho$ are trained individually to obtain a rest pose with additive offsets. $L2$ loss is calculated for each parameter:

$$\mathcal{L}_{shape} = \parallel \beta - \hat{\beta} \parallel^2 \tag{7}$$

$$\mathcal{L}_{exp} = \parallel \psi - \hat{\psi} \parallel^2 \tag{8}$$

$$\mathcal{L}_{rot} = \parallel \rho - \hat{\rho} \parallel^2 \tag{9}$$

where $\hat{\beta}$, $\hat{\psi}$ and $\hat{\rho}$ are the ground-truth for each parameter.

This is the full loss term for training proposed 3D pose estimation from the monocular video method. It is formulated as:

$$\mathcal{L} = \mathcal{L}_{pose} + \mu_1\mathcal{L}_{cam} + \mu_2\mathcal{L}_{shape} + \mu_3\mathcal{L}_{exp} + \mu_4\mathcal{L}_{rot} + \mu_5\mathcal{L}_{tw} + \mu_6\mathcal{L}_{feet} \tag{10}$$

The network minimized the loss calculated to obtain the best estimation. $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ and $\mu_6$ are weights of each loss term and will be learned during training.

## 3.2 Physics Based Optimization

In this section, we discuss the physics-based optimization techniques employed to enhance the stability and ensure realistic behaviour for the estimated 3D poses. The 3D poses are extracted for each frame and connected to form an animation sequence in Blender to be exported to Unity for further optimization.

### 3.2.1 Feet constraint

One of the main issues addressed in this section is the instability of the feet in existing works in the field of 3D human mesh estimation. To tackle this problem, we perform feet adjustment in Unity, using the manually selected plane.

We utilize Unity's *rigidbody* and *mesh collider* for the generated feet meshes. By applying both soft constraints, we can simulate their interaction with the ground plane and ensure proper contact. On top of soft constraints, we implement an additional vertical ray-based constraint. From each ankle and toe joint, two rays are cast downwards with a length equal to an offset value. When these rays hit the ground, it indicates that the feet are in contact with the floor. We then apply a position constraint to the ankle and toe joints, preventing them from moving lower than the plane.

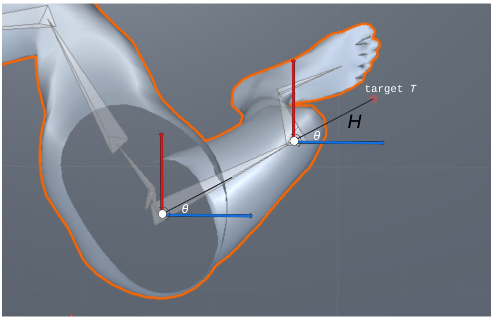


Figure 2: Visualization of Inverse Kinematics target calculation. Target $T$ is used to determine the rotation of the middle joints between the start and end joints when calculating inverse kinematics. The angle for determining the target for knee rotation is the same as the rotation between the root and hip.

The combination of the ray-calculated position constraint and the collision detector effectively prevents feet clipping. For ensuring full body motion smoothness, we employ momentary inverse kinematics (IK). The lower body of the animation will switch to the corresponding inverse kinematics solution instead of its original animation when the feet are about to clip through the floor. Inverse kinematics is a technique that determines the joint positions based on the end joint's position, which in our case is the ankle coordinates. Unity provides the capability to manually adjust the target position $(T)$ that the knee should follow.

To minimize the deviation from the original animation, the coordinates of the target position $(T)$ for the knee are calculated by following:

$$T = (x_{knee} + cos(\theta) \cdot H, y_{knee}, z_{knee} + sin(\theta) \cdot H) \tag{11}$$

where $x_{knee}, y_{knee}$ and $z_{knee}$ represents coordinate of the corresponding knee, $\theta$ is an angle parameter, and $H$ is the distance to target point. The target position is set as shown also in Figure 2. If no clipping is detected, the target position is ignored, and the model follows its original animation.

### 3.2.2 Depth constraint

In addition to the feet constraint, we introduce a soft depth constraint to optimize body jitter caused by an incorrect depth detection from the proposed neural network. We limit the range of motion of the root of the body by a sphere with a diameter equal to twice the body's thickness. This range should be optimised based on the animation and the desired level of constraint. We also allow the constraint to be violated if necessary, which helps to create more natural and smooth movements that resemble real-life motion.

By combining the depth constraint with the previously introduced feet physics optimization, we achieve sufficient movement for estimating GRF. The depth constraint ensures the body moves within a plausible range, while the feet constraints maintain proper contact with the ground. It's worth mentioning that due to the nature of these constraints, they do not completely prevent the feet from floating. In cases where floating feet are observed after the 3D pose estimation process adjusting the plane coordinates may serve as a potential solution, however, this approach should be considered a secondary measure rather than the primary method for addressing the issue.

### 3.3 Ground Reaction Force

As proposed by Newton and Thompson [22], the GRF is equal in magnitude and opposite in direction to the force that the body exerts on the supporting surface through the foot, shown in Figure 3. It is important to note that the friction force can be considered as the third component of GRF. However, measuring friction force is very challenging due to the complexity of the foot-ground interface and the limitations of current force platform technology. As a result, the friction force is often ignored, as reported by [4, 16]

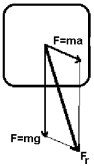


Figure 3: Diagram of basic Force $F_r$ components. It is a joined vector with $F_{ma}$:Acceleration force and $F_{mg}$:Gravity [22]. $F_r$ is used to calculate the ground reaction force on a moving body.

To adapt the GRF calculation to our animation, we compute the force in every frame where the ray and collision detector detects contact. The mass of the body can be the actual mass or assumed based on the height of the subject. To capture a more comprehensive representation of the leg movement and determine the acceleration of the body for GRF accurately, we incorporate the entire lower body and root joint in the calculation. The change in position for acceleration is computed using the following equation:

$$a = N\left(\frac{1}{F}\sum_{F=1}^{F}\left(\frac{root + knee + ankle + toe}{4}\right)\right) \tag{12}$$

where $F$ is the number of frames and $N$ is the normal distribution. By setting $F = 5$, we calculate the mean of the average lower body's joints change in position and normalize the acceleration for every 5 frames. This approach ensures a smooth transition in velocity and provides a more robust estimation of the acceleration.

Finally, combination of the optimized 3D poses, contact detection, and the adapted GRF calculation, allow us to estimate the GRFs for each frame from the original input video.

## 4 Experimental Results

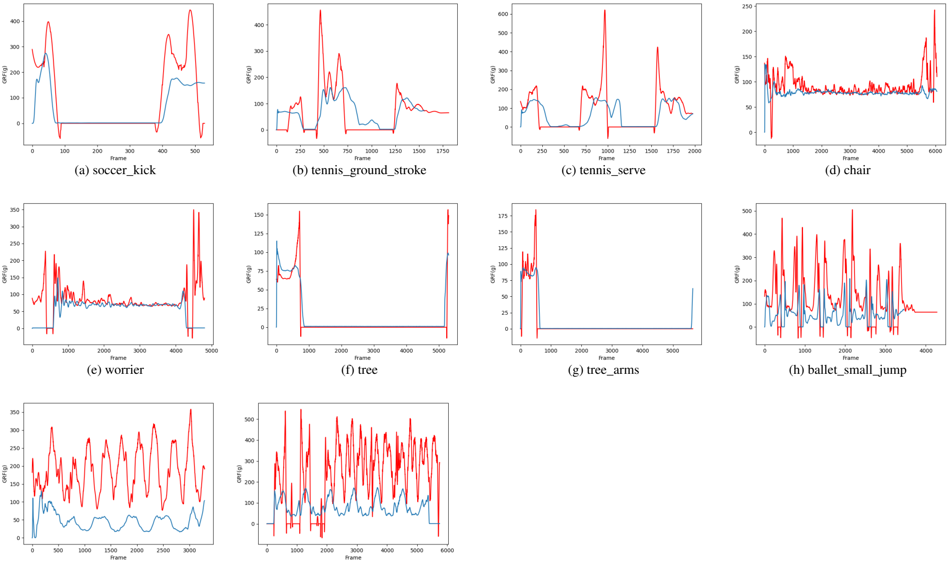
In this section, we will evaluate the performance of our method both quantitatively and qualitatively, against ground truth data from the GroundLink dataset[7].

It is important to note that due to time constraints and limited computational resources, we were unable to conduct a full training of our proposed 3D pose estimation network with the additional loss terms. Instead, we utilized the pretrained weights of the HybrIK-X model[13], which were not trained specifically for this task. We instead applied our physics-based optimization and GRF estimation pipeline to the recovered 3D poses, adapting the pretrained model to our specific use case.

### 4.1 Qualitative Comparison

Overall, the qualitative comparison figures clearly show that our calculated GRFs exhibit similar trends to the ground truth data. The estimated force profiles capture the key characteristics and timing of the GRF patterns, demonstrating the effectiveness of our physics-based optimization and GRF estimation pipeline.

Figure 4 (a) and (b) illustrates the estimated and ground truth GRF curves for a *soccer_kick* and *tennis_ground_stroke* motion. The estimated GRF closely follows the overall shape and timing of the ground truth data. However, the magnitude of the estimated forces



(a) soccer_kick (b) tennis_ground_stroke (c) tennis_serve (d) chair

(e) worrier (f) tree (g) tree_arms (h) ballet_small_jump

(i) dog (j) side_stretch

Figure 4: Calculated ground reaction force (Red) compared against ground truth from GroundLink dataset [7] (Blue) for 10 randomly chosen motions. Horizontal axis: Frame; vertical axis: Ground reaction force (g).

is slightly higher than the ground truth, particularly during the impact phase of the motions, likely caused by the estimation of the weight of the subject.

For steady poses, our method demonstrates a high level of accuracy. Figures 4 (d) and (e) showcase the results of our method applied to continuous *chair* and *warrior* poses, respectively. It is not only showing the accurate estimation of the consistent force profiles during the sustained postures, our method also effectively captures the vertical force transitions during the entering and exiting phases of the poses.

However, some failure cases can be observed especially with motions that contain horizontal weight transitions. Figure 4 (i) and (j) describe the GRF comparison for the *side_stretch* motions and *dog* pose motion. The estimated forces are higher than the ground truth on average for both motions since the force distributed between other points of contact is not captured. In these cases, the slow velocities lead to inaccuracies in the calculation of the velocity term in the GRF equation, resulting in miscalculations of the forces.

## 4.2 Quantitative Comparison

To evaluate the accuracy of our estimated GRFs, we compare them with the ground truth GRF data from the GroundLink dataset. Table 1 presents the mean squared error (MSE) of the calculated GRFs from our proposed approach for both the left and right feet for 10 randomly selected motions. Notably, to the best of our knowledge, this is the first study to utilize a Unity implementation for calculating GRFs. As a result, there is limited availability of directly comparable data from prior works.

The results in Table 1 provide insights into the performance of our method across different human motions, including sports and yoga poses. The MSE values range from 0.003 to 0.150, indicating varying levels of accuracy in the estimated GRFs. Lower MSE values, such as those observed for the *tree_arms* and *tree* motions, suggest a closer match between the estimated and ground truth forces. On the other hand, higher MSE values, like those seen for the *side_stretch* and *dog* motions, indicate larger discrepancies between the estimated and the actual forces. These larger values for the *dog* motions can be explained by the fact that the hands are not considered as points of contact, causing these inaccuracies. Similarly, the *side_stretch* motion is due to the fact that the GRF estimation prioritises front-back velocity over side-to-side motions, making it not very suitable for this approach. Motions with more dynamic and rapid movements, such as the *soccer_kick* and *tennis_serve*, tend to have higher MSE values compared to more static or slow-moving poses, due to the rapid changes in velocity, and direction.

Table 1: Mean Square Error with GroundLink ground truth dataset for 10 randomly chosen motions for both legs. Results are normalized between 0 to 1 by maximum force detected.

| Motions | *Left Leg* | *Right Leg* |
|---|---|---|
| soccer_kick | 0.051 | 0.043 |
| dog | 0.106 | 0.112 |
| chair | 0.013 | 0.028 |
| side_stretch | 0.150 | 0.148 |
| tree_arms | 0.003 | 0.006 |
| tree | 0.004 | 0.011 |
| worrier | 0.017 | 0.009 |
| tennis_serve | 0.023 | 0.005 |
| ballet_small_jump | 0.078 | 0.069 |
| tennis_ground_stroke | 0.029 | 0.016 |

Furthermore, the MSE values for the left and right feet within each motion are generally comparable, suggesting a consistent performance of our method in estimating GRFs for both limbs. However, there are a few cases where the MSE differs slightly between the left and right feet, which could be due to asymmetries in the motion or variations in the quality of the 3D pose estimates.

In summary, our results demonstrate the potential of our proposed approach for non-invasive estimation of GRFs from monocular video on human subjects. The qualitative comparison showcases the ability of our method to capture the key characteristics and trends of the GRF. The quantitative evaluation reveals promising accuracy across various motion types, with room for improvement in handling slow weight transfer poses and refining the force magnitudes. Despite the limitations imposed by the use of pretrained 3D pose estimates, our results highlight the effectiveness of our physics-based optimization and GRF estimation pipeline.

## 5 Conclusion

In this paper, we proposed a pipeline for GRF Estimation from Monocular Video within a healthcare setting. Our approach begins by optimizing feet contact points using newly proposed loss terms (HybrIK-XL), which are then combined with physics-based position constraints to further refine the error in the contact points of the feet. We evaluated our method using the Groundlink dataset [7] by calculating the MSE differences between our estimated GRFs and the ground truth values. The experimental results indicate that our pipeline achieves accurate GRF estimation, with a clear trend of force shifting captured in the estimated profiles, and the proposed pipeline has shown a novel innovation that significantly lowers the difficulty of GRF estimation.

Future work should aim to improve the relationship between the GRF and joint forces, refining the loss term and physics-based optimizations. Furthermore, an ablation study of the optimization step should be performed to confirm its significance. Additionally, evaluation of clinical videos, not feasible within this study due to ethical consent limitations, would confirm the effectiveness of this method within a healthcare setting.

Studies on a purely deep learning method for contact point estimation could lead to a more efficient method of GRF estimation of monocular video without the need for specialized equipment.

## References

[1] Andrea Ancillao, Salvatore Tedesco, John Barton, and Brendan O'Flynn. Indirect measurement of ground reaction forces and moments by means of wearable inertial sensors: A systematic review. *Sensors*, 18(8):2564, Aug 2018. doi: 10.3390/s18082564.

[2] Andrea Ancillao, Salvatore Tedesco, John Barton, and Brendan O'Flynn. Indirect measurement of ground reaction forces and moments by means of wearable inertial sensors: A systematic review. *Sensors*, 18(8), 2018. ISSN 1424-8220. doi: 10.3390/s18082564. URL https://www.mdpi.com/1424-8220/18/8/2564.

[3] Hongsuk Choi, Gyeongsik Moon, Ju Yong Chang, and Kyoung Mu Lee. Beyond static features for temporally consistent 3d human pose and shape from a video. *CoRR*, abs/2011.08627, 2020. URL https://arxiv.org/abs/2011.08627.

[4] M. Damavandi, P. C. Dixon, and D. J. Pearsall. Ground reaction force adaptations during cross-slope walking and running. *Human Movement Science*, 31(1):182–189, 2012. doi: 10.1016/j.humov.2011.06.004.

[5] Kistler Group. Force plates. https://www.kistler.com/en/applications/sensor-technology/biomechanics-and-force-plate/, 2020. Accessed: YYYY-MM-DD.

[6] John K Haas. A history of the unity game engine. 2014.

[7] Xingjian Han, Ben Senderling, Stanley To, Deepak Kumar, Emily Whiting, and Jun Saito. Groundlink: A dataset unifying human body movement and ground reaction dynamics. In *SIGGRAPH Asia 2023 Conference Papers*, SA '23. ACM, December 2023. doi: 10.1145/3610548.3618247. URL http://dx.doi.org/10.1145/3610548.3618247.

[8] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. *CoRR*, abs/1712.06584, 2017. URL http://arxiv.org/abs/1712.06584.

[9] Muhammed Kocabas, Nikos Athanasiou, and Michael J. Black. VIBE: video inference for human body pose and shape estimation. *CoRR*, abs/1912.05656, 2019. URL http://arxiv.org/abs/1912.05656.

[10] Nikos Kolotouros, Georgios Pavlakos, Michael J Black, and Kostas Daniilidis. Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In *ICCV*, 2019.

[11] Marek Kulbacki, Jakub Segen, and Jerzy Paweł Nowacki. 4gait: Synchronized mocap, video, grf and emg datasets: Acquisition, management and applications. *arXiv preprint arXiv:2112.03553*, 2021.

[12] Jiefeng Li, Chao Xu, Zhicun Chen, Bian, Lixin Yang, and Cewu Lu. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. *CoRR*, abs/2011.14672, 2020. URL https://arxiv.org/abs/2011.14672.

[13] Jiefeng Li, Siyuan Bian, Chao Xu, Zhicun Chen, Lixin Yang, and Cewu Lu. Hybrik-x: Hybrid analytical-neural inverse kinematics for whole-body mesh recovery, 2023.

[14] Kevin Lin, Lijuan Wang, and Zicheng Liu. End-to-end human pose and mesh reconstruction with transformers. *CoRR*, abs/2012.09760, 2020. URL https://arxiv.org/abs/2012.09760.

[15] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.

[16] C. Morio, M. J. Lake, N. Gueguen, G. Rao, and L. Baly. The influence of footwear on foot motion during walking and running. *Journal of Biomechanics*, 42(13):2081–2088, 2009. doi: 10.1016/j.jbiomech.2009.06.015.

[17] Lucas Mourot, Ludovic Hoyet, François Le Clerc, and Pierre Hellier. Underpressure: Deep learning for foot contact detection, ground reaction force estimation and footskate cleanup, 2022.

[18] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[19] Davis Rempe, Leonidas J. Guibas, Aaron Hertzmann, Bryan Russell, Ruben Villegas, and Jimei Yang. Contact and human dynamics from monocular video. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[20] Erfan Shahabpoor and Aleksandar Pavic. Measurement of walking ground reactions in real-life environments: A systematic review of techniques and technologies. *Sensors*, 17(9), 2017. ISSN 1424-8220. doi: 10.3390/s17092085. URL https://www.mdpi.com/1424-8220/17/9/2085.

[21] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5693–5703, 2019.

[22] D. Thompson. Ground reaction force. University of Oklahoma Health Sciences Center, April 3 2002. URL https://ouhsc.edu/bserdac/dthompso/web/gait/kinetics/GRFBKGND.HTM. Accessed: 2024-04-03.

[23] Shashank Tripathi, Lea Müller, Chun-Hao P. Huang, Omid Taheri, Michael J. Black, and Dimitrios Tzionas. 3d human pose estimation via intuitive physics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4713–4725, June 2023.

[24] Kevin Xie, Tingwu Wang, Umar Iqbal, Yunrong Guo, Sanja Fidler, and Florian Shkurti. Physics-based human motion estimation and synthesis from videos. *CoRR*, abs/2109.09913, 2021. URL https://arxiv.org/abs/2109.09913.

[25] Guanyu Yang, Wen-Hao Hsu, Kevin Chou, Jiajun Hu, Jiajun Wu, Daniel Yamins, and Taku Komura. From image to stability: Learning dynamics from human pose. *arXiv preprint arXiv:2109.14076*, 2021.

[26] Xinyu Yi, Yuxiao Zhou, Marc Habermann, Soshi Shimada, Vladislav Golyanik, Christian Theobalt, and Feng Xu. Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors, 2022.

[27] Ye Yuan, Jiaming Song, Umar Iqbal, Arash Vahdat, and Jan Kautz. Physdiff: Physics-guided human motion diffusion model, 2023.

[28] Yeqing Zhu, Di Xia, and Heng Zhang. Using wearable sensors to estimate vertical ground reaction force based on a transformer. *Applied Sciences*, 13(4), 2023. ISSN 2076-3417. doi: 10.3390/app13042136. URL https://www.mdpi.com/2076-3417/13/4/2136.