Weakly-supervised Learning of Human Dynamics - Appendix

Petrissa Zell[1], Bodo Rosenhahn[1], and Bastian Wandt[1]

Leibniz University Hannover, 30167 Hannover, Germany
zell@tnt.uni-hannover.de

A Weight gradients of dynamics layers

For the forward layer, the gradients of simulated states $\boldsymbol{x}$ with respect to the layer input $\boldsymbol{p}_f$ can be calculated using sensitivity analysis: The partial differentiation of the equations of motion (Eq. (8)) by $\boldsymbol{p}_f$ yields a second differential equation,

$$\frac{d}{dt}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}_f} = \frac{\partial}{\partial \boldsymbol{p}_f}\left(\begin{bmatrix}\dot{\boldsymbol{q}}\\ \boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{F}}\end{bmatrix} \odot \boldsymbol{d}\right), \tag{A1}$$

for the gradients $\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}}$. The integration of this differential equation supplies the desired gradients for the backpropagation through the forward layer.

Therefore, the forward pass through the forward layer includes the numerical integration of equations (8) and (A1) with storage of the resulting gradients. During the backward pass the gradient of the loss is propagated by multiplication with the stored gradients according to the chain rule:

$$\frac{\partial L_{\text{forward}}}{\partial \boldsymbol{p}_f} = \sum_{t=1}^{n} \frac{\partial L_{\text{forward}}}{\partial \boldsymbol{x}_t^{sim}} \frac{\partial \boldsymbol{x}_t^{sim}}{\partial \boldsymbol{p}_f}. \tag{A2}$$

The calculation of gradients for the inverse layer is straight forward, since there is no dependency between the resulting residual forces and moments of different time frames. Given the input $\boldsymbol{p}_i = [\boldsymbol{l}_{\text{sub}}, \boldsymbol{F}_{c_t}]$, the gradients can be calculated by

$$\begin{aligned}\frac{\partial L_{\text{inverse}}}{\partial \boldsymbol{p}_i} = \frac{4}{n}\sum_{t=1}^{n}\Bigg[&\|\boldsymbol{F}_{\text{res}_t}\|_2^{-1}\left(\frac{\partial \boldsymbol{F}_{\text{res}_t}}{\partial \boldsymbol{l}_{\text{sub}}} + \frac{\partial \boldsymbol{F}_{\text{res}_t}}{\partial \boldsymbol{F}_{c_t}}\right)\\ &+\|\boldsymbol{M}_{\text{res}_t}\|_2^{-1}\left(\frac{\partial \boldsymbol{M}_{\text{res}_t}}{\partial \boldsymbol{l}_{\text{sub}}} + \frac{\partial \boldsymbol{M}_{\text{res}_t}}{\partial \boldsymbol{F}_{c_t}}\right)\Bigg].\end{aligned} \tag{A3}$$

B Demographic information

The recorded data set encompasses 195 walking and 75 running sequences executed by 22 healthy subjects. Demographic information can be found in Table B1. All subjects volunteered to participate in the study and signed an informed consent form. The study is part of the Individualized Implant Placement project funded by the European Research Council (ERC-2013-PoC) and was approved by the ethics commission of the Hannover Medical School (MHH). In order to increase and balance our data set, we augment by mirroring the kinematics and dynamics at the sagittal plane [1].

Table B1. Demographic table of participating subjects.

subject ID	gender	height	weight	BMI
1	m	1.78	93.5	30
2	m	1.94	88.8	24
3	m	1.86	68.3	20
4	f	1.71	66.6	23
5	m	1.80	68.3	21
6	f	1.73	55.7	19
7	f	1.69	65.5	23
8	m	1.71	61.8	21
9	m	1.81	67.9	21
10	m	1.88	74.4	21
11	m	1.81	79.3	24
12	m	1.85	74.5	22
13	m	1.67	83.8	30
14	m	1.85	95.8	28
15	m	1.84	68.8	20
16	m	1.75	81.4	27
17	m	1.72	79.4	27
18	f	1.70	68.0	24
19	m	1.80	72.4	22
20	f	1.74	70.5	23
21	m	1.80	83.5	26
22	m	1.79	69.9	22

C Application to CMU data

In order to test the generalizability to a different 3D motion data set, we apply our method to walking sequences, taken from the CMU data base [1]. Figure 1 shows an exemplary comparison between the baseline method and F-net. Since

[1] The sagittal plane is spanned by the vertical axis and the direction of movement.

the baseline network can only be trained in a supervised manner, it is exclusively trained on our own data set. F-net is additionally trained on CMU-samples using the forward-loss. The result indicates, that the additional cyclic training helps to bridge the domain gap between the two sets: In contrast to the baseline, F-net yields symmetric vertical GRF for both feet and the forces stay closer to zero during frames without contact.

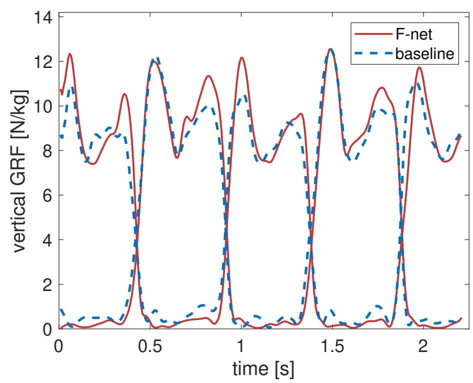

Fig. 1. Regressed vertical GRF for an example sequence taken from the CMU data base.

D Noise experiment

Zero mean gaussian noise with standard deviation σ is added to the joint angles of the motion training set and the test sequences. The results are presented in Table B2. In the case of cFI-net we simulate noisy contact detections in addition to the angle noise. For this purpose, 10 % of the contact labels are randomly chosen and switched with a probability of 50 %.

Angle noise of $\sigma = 0.3$ deg only marginally effects the performance of both networks. Overall, F-net is slightly more robust against noise than cFI-net, in this experiment. The inverse layer matches GRF/M to the noisy motion, resulting in a larger difference between inferred forces and ground truth. In contrast to that, F-net primarily learns GRF/M from the ground truth set. Concerning the JT, the robustness of both methods is comparable. During training, the JT output of both networks is controlled by the MSE and the forward-loss in a similar way.

Table B2. Influence of angle noise: RMSE ϵ and rRMSE ϵ_r of GRF/M and JT regression results for the gait data set with noisy motion training set and noisy test sequences.

method	σ [deg]	ϵ_f [N/kg]	ϵ_{r_f} [%]	ϵ_m [Nm/kg]	ϵ_{r_m} [%]	ϵ_τ [Nm/kg]
F-net	0.3	0.626	14.9	0.058	22.4	0.062
	0.6	0.832	16.8	0.061	22.7	0.068
	1.1	0.898	17.8	0.066	24.3	0.073
	2.3	1.096	19.7	0.073	25.4	0.092
cFI-net	0.3	0.674	13.8	0.062	23.4	0.060
	0.6	0.908	16.7	0.088	26.4	0.076
	1.1	0.998	17.5	0.082	25.9	0.087
	2.3	1.143	19.5	0.079	25.7	0.096

References

1. CMU: Human motion capture database (2014), http://mocap.cs.cmu.edu/