From Image to Stability: Learning Dynamics from Human Pose

Jesse Scott*[1], Bharadwaj Ravichandran*[1], Christopher Funk[2], Robert T. Collins[1], Yanxi Liu[1]

[1]School of EECS, Pennsylvania State University
[2]Kitware, Inc.
(jus121,bzr49)@psu.edu, christopher.funk@kitware.com,
(rcollins, yanxi)@cse.psu.edu

Abstract. We propose and validate two end-to-end deep learning architectures to learn foot pressure distribution maps (dynamics) from 2D or 3D human pose (kinematics). The networks are trained using 1.36 million synchronized pose+pressure data pairs from 10 subjects performing multiple takes of a 5-minute long choreographed Taiji sequence. Using leave-one-subject-out cross validation, we demonstrate reliable and repeatable foot pressure prediction, setting the first baseline for solving a non-obvious pose to pressure cross-modality mapping problem in computer vision. Furthermore, we compute and quantitatively validate Center of Pressure (CoP) and Base of Support (BoS), two key components for stability analysis, from the predicted foot pressure distributions.

Keywords: Stability; Center of Pressure; Base of Support; Foot Pressure Estimation; 3D Human Pose, Deep Regression Models

1 Introduction

Current computer vision research on human pose focuses on extracting skeletal kinematics from images or video [9, 12, 15, 16, 20, 24, 47, 68]. A more effective analysis of human movement should take into account the dynamics of the human body [59]. Understanding body dynamics such as foot pressure is essential to study the effects of perturbations caused by external forces and torques on the human postural system, which change body equilibrium in static posture and during locomotion [73]. Computing stability from visual data could unlock a wide range of applications in the fields of healthcare, kinesiology, and robotics.

For understanding the relation between a body pose and the corresponding foot pressure of a human subject (Figure 1), we explore two deep convolutional residual architectures, PressNet and PressNet-Simple (Figure 5), and train them on a dataset containing 1,363,400 data pairs of body pose with corresponding foot pressure measurements. Body pose is input to a network as either 2D or 3D human joint locations extracted from the Openpose [12] Body25 model (3D

[*]Co-First Authors

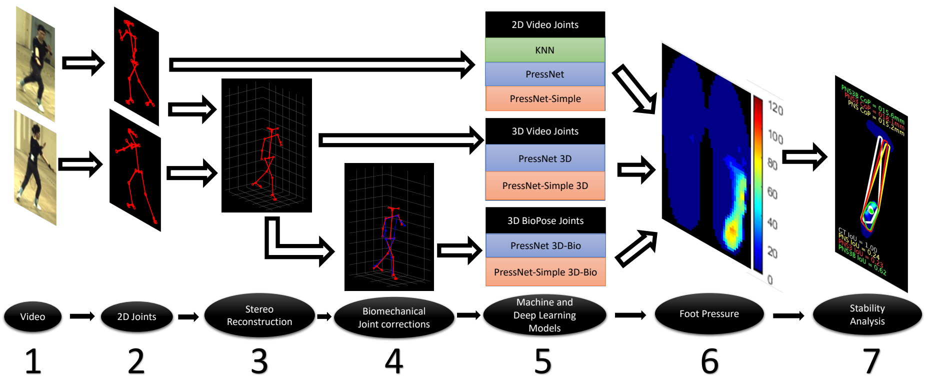
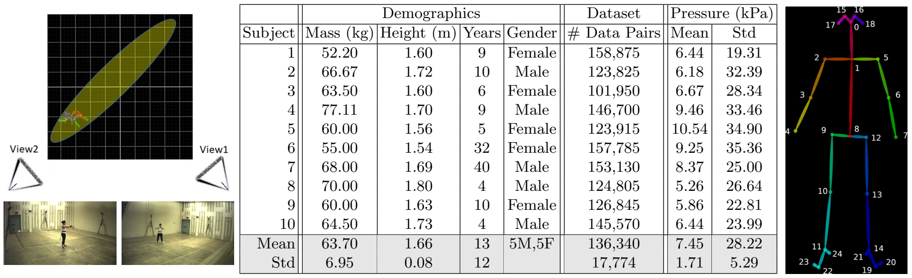


Fig. 1: Our PressNet and PressNet-Simple networks learn to predict a foot pressure heatmap from 2D or 3D human body joints. We also compute Center of Pressure (CoP) and Base of Support (BoS), two key components for stability estimation, from predicted foot pressure distributions (rightmost).


View2
View1
15
16
17
0
18
2
1
5
3
6
4
9
8
12
7
10
13
11
24
14
21
23
22
19
20

	Demographics				Dataset	Pressure (kPa)	
Subject	Mass (kg)	Height (m)	Years	Gender	# Data Pairs	Mean	Std
1	52.20	1.60	9	Female	158,875	6.44	19.31
2	66.67	1.72	10	Male	123,825	6.18	32.39
3	63.50	1.60	6	Female	101,950	6.67	28.34
4	77.11	1.70	9	Male	146,700	9.46	33.46
5	60.00	1.56	5	Female	123,915	10.54	34.90
6	55.00	1.54	32	Female	157,785	9.25	35.36
7	68.00	1.69	40	Male	153,130	8.37	25.00
8	70.00	1.80	4	Male	124,805	5.26	26.64
9	60.00	1.63	10	Female	126,845	5.86	22.81
10	64.50	1.73	4	Male	145,570	6.44	23.99
Mean	63.70	1.66	13	5M,5F	136,340	7.45	28.22
Std	6.95	0.08	12		17,774	1.71	5.29

(A) Two Views (B) Subject Demographics (C) Joints

Fig. 2: **(A)** Top-down view of the motion capture space highlighting the region of performance relative to the two video cameras. **(B)** Dataset statistics. A total of 1,363,400 frames of data have been collected, providing **(C)** 25 body joints from each video frame, time-synchronized with foot pressure map data.

joints are derived by 2-view stereo triangulation of 2D joints). Each network predicts a foot pressure heatmap as output, providing an estimated distribution of pressure applied at different foot locations (Figure 1, stage 6).

The main contributions of this work include **1) Novelty**: Our PressNet and PressNet-Simple networks are the first vision-based networks to regress foot pressure (dynamics) from 3D or 2D body pose (kinematics). Furthermore, we introduce a 3D Pose Estimation Network (BioPose) that enhances body joint positions biomechanically. **2) Dataset**: We have collected the largest synchronized Video, motion capture (MoCap), and foot pressure dataset of a 5-minute long, complex human movement sequence (Figure 2). **3) Application**: For validation, Center of Pressure (CoP) and Base of Support (BoS), two key components in the analysis of stability, are bench-marked for potential future applications.

2 Related Work

Seethapathi et al. [59] reviewed the limitations of video-based measurement of human motion for use in movement science, and indicated that more accurate kinematics and estimation of dynamics information, such as contact forces, should be a key research goal in order to use computer vision as a tool in biomechanics. In this paper, we use body kinematics to predict foot pressure dynamics and to develop a quantitative method to analyze human stability from video. Earlier work in computer vision and graphics has incorporated dynamics equations into models of human motion and person tracking, and has even estimated contact forces from video and kinematics [7, 8, 39, 42, 71], but their estimates of contact dynamics tend to be simple force vectors rather than full foot pressure maps, as in our work.

Studying human stability during standing and locomotion [3, 19, 38] is typically addressed by direct measurement of foot pressure using force plates or insole foot pressure sensors. Previous studies have shown that foot pressure patterns can be used to discriminate between walking subjects [50, 70]. Instability of the CoP of a standing person is an indication of postural sway and thus a measure of a person's ability to maintain balance [27, 28, 37, 49]. Grimm et al. [23] predicts the pose of a patient using foot pressure mats. The authors of [45] and [55] evaluate foot pressure patterns of 1,000 subjects ages 3 to 101 and determine there is a significant difference in the contact area by gender but not in magnitude of foot pressure for adults. As a result, the force applied by females is lower but is accounted for by female mass also being significantly lower. In [11], a depth regularization model is trained to estimate dynamics of hand movement from 2D joints obtained from RGB video cameras. Stability analysis of 3D printed models is presented in [4, 53, 54]. Although these are some insightful ways to analyze stability, there has been no vision-based or deep learning approach to tackle this problem.

Estimation of 2D body pose in images is a well-studied problem in computer vision, with state-of-the-art methods being based on deep networks [9, 12, 15, 16, 20, 22, 24, 30, 47, 66, 68]. We adopt one of the more popular approaches, CMU's OpenPose [12], to compute the 2D pose input to our networks. Success in 2D human pose estimation also has encouraged researchers to detect 3D skeletons by extending existing 2D human pose detectors [6, 14, 44, 46, 48, 62, 75] or by directly using image features [1, 51, 57, 64, 74]. Martinez et al. [44] showed that given high-quality 2D joint information, the process of lifting 2D pose to 3D pose can be done efficiently using a relatively simple deep feed-forward network. All

Table 1: Comparison of our dataset with other available human pose datasets.

Name	# Data samples	# Subjects	Scenario	MoCap joints	Image joints	Foot pressure	Humans per image
Human3.6M [32, 33]	3,600,000	11	Indoor	✓	✓	-	Single
HumanEva [61]	80,000	4	Indoor	✓	✓	-	Single
MPII Human Pose [2]	25,000	N/A	Indoor/Outdoor	-	✓	-	Single/Multiple
MS-COCO [40]	200,000	N/A	Indoor/Outdoor	-	✓	-	Multiple
PoseTrack [34]	66,000	N/A	Indoor/Outdoor	-	✓	-	Multiple
Taiji Stability (Ours)	1,363,400	10	Indoor	✓	✓	✓	Single

these papers concentrate on pose estimation by learning to infer joint angles or joint locations, which can be broadly classified as learning basic kinematics of a body skeleton. These methods do not predict external torques/forces exerted by the environment, balance, or physical interaction of the body within the scene.

Table 1 shows a summary of available pose datasets. Human3.6M dataset [32,33] is one of the top public datasets that is used for pose estimation tasks. It has 3.6 million frames of synchronized 3D MoCap-based human pose and video data. This data was collected from 6 male and 5 female subjects for 17 different scenarios. HumanEva [61] is a similar dataset with MoCap and video data, but it is smaller in size than the Human3.6M dataset. Another dataset that is widely popular is the MPII Human Pose dataset [2]. This dataset consists of 25,000 images containing over 40,000 individuals with Ground Truth (GT) human body joints, covering over 410 human activities. The Posetrack [34] dataset consists of about 1,400 video sequences with over 66,000 annotated video frames and 276,000 body pose annotations. The MS-COCO [40] dataset has more than 200,000 images and 250,000 individuals with labeled keypoints of human pose.

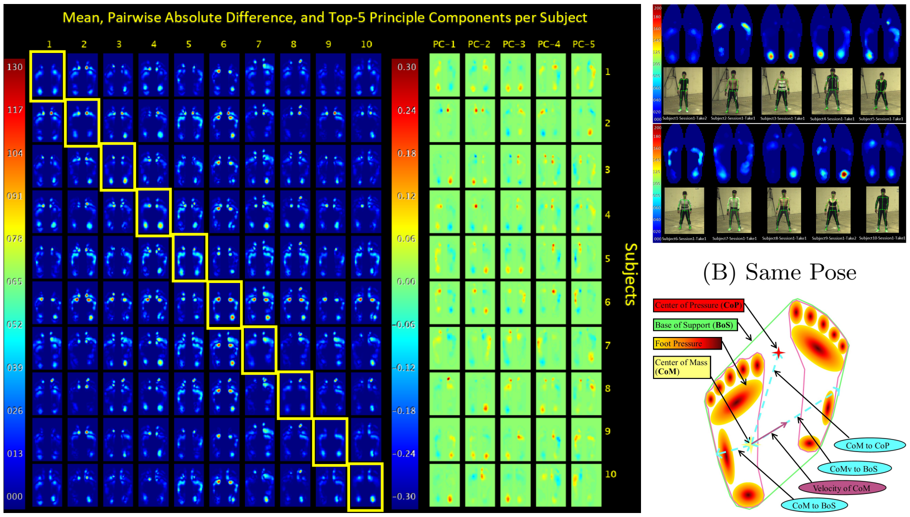
A major motivation for computing foot pressure maps from pose is to estimate body stability from video in the wild, accurately and economically, rather than in a biomechanics lab. Fundamental elements used in stability analysis (Figure 3C) include Center of Mass (CoM), Base of Support (BoS), and Center of Pressure (CoP). The relative locations of CoP, BoS, and CoM have been identified as a determinant of stability in a variety of tasks [27,28,49].

3 Our Approach and Motivation

Mapping from human pose to foot pressure (Figure 1) is an ill-posed problem. On the one hand, similar poses of different subjects can yield different foot pressure maps (Figure 3B) due to differences in movement, mass, height, gender, and foot shape. On the other hand, PCA analysis (Figure 3A) suggests the top principal components capture statistically similar "modes" of variation across subjects. Thus, we formulate our problem as learning foot pressure distribution conditioned on human pose rather than trying to directly regress precise foot pressure magnitude. For simplicity, we assume the conditional distribution of pressure given pose is Gaussian, with a mean that can be learned through deep learning regression using MSE and KL Divergence loss. Our networks are trained to map from pose, encoded as 25 joint locations (2D or 3D), to the mean of a corresponding foot pressure map intensity distribution (Figure 5).

3.1 Data Collection and Pre-Processing

We have collected a tri-modal dataset containing synchronized video, motion capture, and foot pressure data (Figure 2) of 24-form simplified *Taiji Quan* (Taiji or Tai Chi) [72]. Justifications for this choice include 1) that Taiji is a low-cost, low-impact, slow, and hands-free movement sequence, aiming at enhanced balance; meanwhile, it contains ordinary body poses and movements such as



(B) Same Pose

(A) Mean Pressure Comparison

(C) Key Stability Terms

Fig. 3: **(A)** Left: Pairwise absolute differences between mean foot pressure across all subjects with inter-subject comparison of differences in pressure magnitude and spatial distribution. Mean pressure is provided on the diagonal (yellow boxes). Right: Top-5 Principal Components of foot pressure data per subject. **(B)** Same "starting pose" yields different foot pressure for different subjects. **(C)** Basic concepts in stability analysis, including Center of Pressure, Center of Mass, and Base of Support.

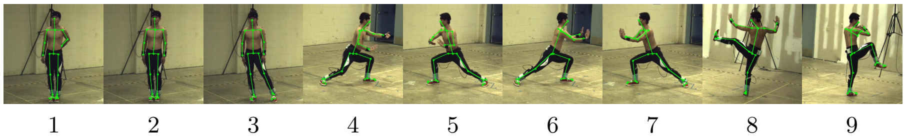


Fig. 4: Example Taiji poses similar to ordinary movements: 1 - standing with hand behind, 2 - standing with two arms down, 3 - step to left, 4 - bump (arm) to left, 5 - bump (arm) to right, 6 - push to left, 7 - push to right, 8 - left kick, and 9 - right kick.

stand, turn, pull, push, bump, and kick (Figure 4) [72]; 2) Simplified 24-form Taiji is practiced worldwide by people of every gender, race and ages; 3) the Taiji routine (5 min) is significantly longer than existing publicly available motion capture (MoCap) sequences in the computer vision community (Section 2).

Pose Extraction: Synchronized video is collected at 50 fps from two Vicon Vue cameras. Locations for 2D body joints are first estimated in each video frame using the OpenPose Body25 model, which uses non-parametric representations called Part Affinity Fields to regress joint positions and body segment connections between the joints [12]. The output from OpenPose has X, Y pixel coordinates and confidence of prediction for each of the 25 modeled joints. To generate 3D joint locations, a confidence-weighted stereo triangulation is performed on

Table 2: **(A)** Subject-wise and **(B)** Joint-wise L2 distance error of 3D pose data. The mean, std, min, median, and max are provided (in mm) for both OpenPose and BioPose joint locations as compared to motion capture joint data. Difference shows percentage improvement of BioPose over OpenPose.

Subject-wise L2 error relative to Motion Capture									
Sub	OpenPose(mm)			BioPose(mm)			Difference(%)		
#	mean	std	med	mean	std	med	mean	std	med
1	52.6	54.1	43.3	33.7	27.5	28.4	35.9	49.2	34.4
2	54.3	62.5	42.3	35.9	33.7	29.5	33.9	46.2	30.3
3	54.2	49.9	43.5	34.4	27.4	29.5	36.5	45.0	32.2
4	57.6	55.1	46.4	33.8	22.9	30.0	41.4	**58.5**	35.3
5	54.7	78.6	42.3	37.9	68.4	28.4	30.8	13.0	33.0
6	51.9	47.4	43.0	33.0	24.5	28.5	36.5	48.3	33.8
7	53.4	46.7	45.3	30.8	21.8	26.4	42.2	53.4	41.8
8	51.3	46.7	43.2	30.4	21.2	26.3	40.7	54.6	39.2
9	55.5	51.6	45.8	31.4	30.5	25.5	**43.3**	41.0	**44.3**
10	50.9	51.4	42.2	31.0	24.8	26.0	39.1	51.7	38.5
Mean	53.6	54.4	43.7	33.2	30.3	27.8	38.0	44.4	36.3
Std	2.1	9.8	1.5	2.4	14.0	1.7	4.0	12.7	4.5

(A) Subject-wise

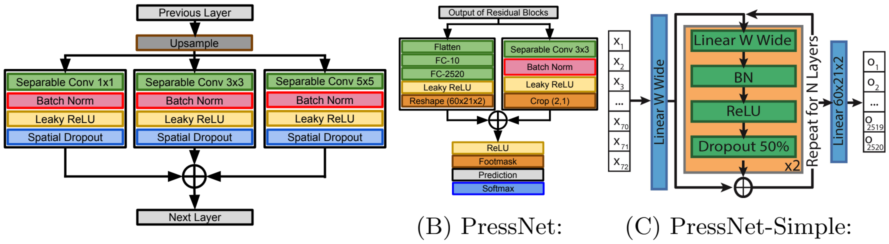
Joint-wise L2 error relative to Motion Capture									
Joint	OpenPose(mm)			BioPose(mm)			Difference(%)		
Location	mean	std	med	mean	std	med	mean	std	med
Rshoulder	34.7	18.7	33.0	26.4	16.6	24.7	24.0	11.3	25.3
Relbow	52.0	49.7	41.1	34.1	28.3	27.8	34.5	43.0	32.2
Rwrist	67.9	85.8	44.9	42.2	40.5	31.9	37.8	52.8	29.1
Lshoulder	40.9	21.6	39.3	28.2	17.0	26.4	31.2	21.2	32.7
Lelbow	65.5	60.5	46.7	37.9	29.9	31.0	42.1	50.5	33.5
Lwrist	87.9	100.5	51.1	49.5	43.8	36.7	43.7	**56.4**	28.3
Rhip	55.3	22.7	53.3	34.9	18.5	33.9	36.9	18.5	36.3
Rknee	51.3	34.5	48.9	28.2	24.3	25.0	45.2	29.6	48.8
Rankle	49.7	47.8	44.2	26.9	31.5	22.3	46.0	34.0	**49.6**
Lhip	59.9	24.9	59.3	32.0	18.0	30.8	**46.5**	27.5	48.1
Lknee	39.7	28.6	37.1	32.4	25.6	29.7	18.3	10.6	20.0
Lankle	42.1	41.2	37.6	27.8	32.6	23.2	33.9	21.0	38.2
Mean	53.6	54.4	43.7	33.2	30.3	27.8	38.0	44.4	36.3

(B) Joint-wise

2D Openpose joints across the two synchronized and calibrated camera views. Finally, the 3D joints are corrected spatially using a deep regression network, named BioPose, trained separately to predict offsets between triangulated OpenPose joints and biomechanical joints computed from motion capture data using the Vicon Plug-in-Gait module (Tables 2A and 2B). Pose detectors and BioPose corrections were tested and evaluated in detail by [56]; showing that OpenPose is more biomechanically accurate than HRNet [63] and Biopose correction of OpenPose creates the most biomechanical accuracy joint locations.

Foot Pressure: Foot pressure is collected at 100 fps using a Tekscan F-Scan insole pressure measurement system. Each subject wears a pair of canvas shoes with cut-to-fit capacitive pressure measurement insoles. Maximum recorded pressure values are clipped at an upper bound of 862 kPa based on the technical limits of the pressure measurement sensors. The foot pressure heatmaps are 2-channel images of size 60 × 21 (Figure 1) and have been evaluated as accurate measurement sensors by [29].

Dataset Statistics: Figure 2B presents demographic information of the 10 subjects. Each subject performs two to three sessions of 24-form Taiji at an average of four repeated performances per session. The dataset contains a total of 1,363,400 frames of synchronized video body pose and foot pressure maps. This new dataset captures significant statistical variations: 1) Diversity in the subjects in terms of gender, age, mass, height, and years of experience in Taiji practice for amateurs and professionals. 2) Kernel density plots (on project page) of the distributions of body joint locations show the subject performances are statistically similar to one another spatially. 3) PCA analysis (Figure 3) of foot pressure highlights that each subject has a unique pressure distribution relative



(A) PressNet: Residual Block

(B) PressNet: Final Layers

(C) PressNet-Simple: Complete Network

Fig. 5: Our foot pressure regression architectures have a 96-coordinate input representing 24 3D joint locations and confidences ($24 \times 4 = 96$) and a 2520-prexel output representing 60×21 pressure maps for both feet ($60 \times 21 \times 2 = 2520$). **(A)** A residual block, one of the building blocks of PressNet network, upsamples the data and computes features. **(B)** Final set of layers of PressNet include a fully connected layer and a concurrent branch to preserve spatial consistency. **(C)** The PressNet-Simple network architecture is defined by two hyperparameters: the depth (# of layers, N) of the network and the width (# of fully connected nodes, W) of those layers.

to other subjects, but the top principal components encode similar modes of variation (e.g., variability in left/right foot pressure, toe/heel pressure, etc.).

Preprocessing: Body joints are centered by subtracting off the hip center joint location (making it the origin) to remove camera-specific offsets during video recording. Other joint locations are normalized per body joint by subtracting each dimension (2D or 3D) by the mean and dividing by its standard deviation, leading to a zero-mean, unit-variance distribution.

Foot pressure data is recorded in kilopascals (kPa) at discretized sensor locations (prexels) on the shoe insoles. Prexel values are clipped between 0 to 862 kPa based on pressure sensor technology limitations. The clipped data is further normalized by dividing each prexel by its max intensity value over the entire training set. The left and right normalized foot pressure maps are concatenated as two channels to form a ground truth heatmap of size ($60 \times 21 \times 2$), with prexel intensities in the range $[0, 1]$.

3.2 PressNet Network

The design of our PressNet network (Figures 5A and 5B) is initially motivated by the residual generator of the Improved Wasserstein GAN [25]. We use a generator-inspired architecture because our input is 1D body joints and the output is a 2D foot pressure heatmap. This design aids in capturing information at different resolutions, acting like a decoder network for feature extraction. The primary aim of this network is to extract features without loss of spatial information across different scales.

PressNet is a feed forward Convolutional Neural Network with an input layer that is a flattened vector of joint coordinates of size 96×1 (24 joints × 4, consisting of x,y,z coordinates and joint detection confidences). The input is processed through a fully connected layer with an output dimension of 6144 × 1. This output is reshaped into an image of size 4 × 3 with 512 channels. The network contains four residual convolution blocks that perform nearest neighbor upsampling. Each residual block of PressNet (Figure 5A) has three parallel convolution layers with kernel sizes 5 × 5, 3 × 3 and 1 × 1. The number of channels of each residual block is progressively halved as the resolution doubles, starting at 512 channels and decreasing to 64. The output of the final residual block is split and sent to a convolutional branch and a fully connected branch (Figure 5B). The convolutional branch serves to preserve spatial coherence while the fully connected branch has a field of view over the entire prediction. PressNet contains separable convolutions [17], batch normalization (BN) [31], spatial dropouts [67] and leaky ReLU [43] layers.

3.3 PressNet-Simple Network

The "simple yet effective" network of [44] was originally designed to jointly estimate the unobserved third dimension of a set of 2D body joint coordinates (pose) on a per frame basis. We use it as a basis for our PressNet-Simple architecture (Figure 5C) by adapting the network architecture to use a modified pose input format and by completely reconfiguring the output format to produce pressure map data of each foot. The input pose coordinates are passed through a fully connected layer then through a sequence of N repeated layers. Each of the N layers has two iterations of the sequence: fully connected, batch normalization, ReLU, and 50% dropout layers. The result of each of the N layer sequences is then added to the results from the previous layer sequence (N-1) and finally passed through a 2520 fully connected layer to produce the output foot pressure. The PressNet-Simple architecture is configured via two hyper-parameters: the depth (# of layers, N) of the network and the width (# of fully connected nodes, W) of those layers. For this study, through empirical testing, it was determined that the optimal hyper-parameters are N=4 and W=2560. Because of the sequential nature of this network with fully connected layers, this network architecture does not maintain the spatial coherence that PressNet has established with upsampling and convolutional layers.

3.4 Training Details

We use a Leave-One-Subject-Out (LOSO) cross-validation to determine how the network generalizes to an unseen individual. Furthermore, the training data is split sequentially in a 9:1 ratio where the smaller split is used as validation data.

PressNet is trained for 35 epochs with a piecewise learning rate starting at 10^{-4} and a batch size of 32 and takes 7.5 hours to train each LOSO data split on a NVIDIA Tesla P100 GPU. A binary footmask (produced by the foot pressure capturing system) is element-wise multiplied with the predictions of the network. This enables the network to not have to learn the approximate shape of the foot in the course of training, only the pressure distribution. The learning rate is reduced by 50% after every 13 epochs to ensure a decrease in validation loss

Table 3: Analysis for each network architecture by subject using error metrics: Mean Absolute Error (MAE), Similarity (SIM), KL Divergence (KLD), and Information Gain (IG). Results from inference on input poses with 25 valid joints are included in the evaluation. Best values are shown in bold. Arrow indicates direction of better value. Networks Evaluated are KNN 2D with K=5 (K5), PressNet-Simple 2D (PNS2), PressNet-Simple 3D (PNS3), PressNet-Simple 3D using BioPose(PNS3B), and PressNet 3D using BioPose and KL loss (PN3BK).

Mean Error of Estimated Foot Pressure relative to Measured Ground Truth																				
Sub #	Mean Absolute Error (MAE) ↓					Similarity (SIM) ↑					KL Divergence (KLD) ↓					Information Gain (IG) ↑				
1	9.56	7.55	7.21	**7.06**	N/A	0.30	0.44	0.49	**0.50**	0.48	17.02	2.79	2.04	2.65	**1.16**	-4.35	-0.74	-0.50	-0.77	**-0.27**
2	9.98	**9.44**	10.10	9.90	N/A	0.22	0.23	0.25	**0.28**	0.27	14.34	5.21	4.39	3.50	**1.98**	-1.23	-0.41	-0.34	-0.27	**-0.14**
3	9.73	8.82	**7.77**	7.87	N/A	0.31	0.39	0.44	**0.44**	0.43	14.80	2.95	1.61	2.02	**0.43**	-2.76	-0.54	-0.29	-0.39	**-0.18**
4	10.91	9.38	**9.12**	10.07	N/A	0.32	0.45	**0.48**	0.42	0.43	14.42	1.88	1.59	3.53	**0.43**	-3.70	-0.52	-0.42	-0.92	**-0.23**
5	11.21	10.14	9.28	**9.22**	N/A	0.40	0.47	**0.55**	0.53	0.50	12.19	1.94	1.51	1.96	**0.50**	-2.65	-0.44	-0.36	-0.46	**-0.21**
6	11.06	9.98	10.04	**9.13**	N/A	0.38	0.45	0.45	**0.51**	0.44	10.48	1.54	1.07	1.60	**0.44**	-1.56	-0.22	-0.14	-0.25	**-0.12**
7	11.08	10.18	**8.97**	9.14	N/A	0.26	0.34	0.42	**0.42**	0.39	18.49	4.44	2.70	3.19	**1.87**	-4.39	-0.98	-0.65	-0.74	**-0.37**
8	8.94	8.31	**7.32**	7.75	N/A	0.30	0.33	0.34	**0.38**	0.38	13.32	3.32	3.02	2.48	**1.62**	-1.51	-0.34	-0.32	-0.29	**-0.17**
9	9.26	8.24	**7.43**	7.63	N/A	0.32	0.37	**0.47**	0.45	0.43	13.09	2.82	1.45	2.04	**1.14**	-2.16	-0.46	-0.25	-0.35	**-0.16**
10	8.82	7.44	7.53	**7.26**	N/A	0.33	**0.44**	0.42	0.41	0.41	15.30	3.18	3.14	3.81	**1.50**	-2.84	-0.60	-0.65	-0.73	**-0.24**
Mean	10.06	8.95	**8.48**	8.50	N/A	0.31	0.39	0.43	**0.43**	0.42	14.35	3.01	2.25	2.68	**1.11**	-2.71	-0.53	-0.39	-0.52	**-0.21**
Std	**0.89**	0.98	1.09	1.05	N/A	**0.05**	0.07	0.08	0.07	0.06	2.18	1.08	0.98	0.74	**0.59**	1.09	0.20	0.16	0.24	**0.07**
Model	K5	PNS2	PNS3	PNS3B	PN3BK	K5	PNS2	PNS3	PNS3B	PN3BK	K5	PNS2	PNS3	PNS3B	PN3BK	K5	PNS2	PNS3	PNS3B	PN3BK

with training. KL Divergence (KL) is used as the loss function along with Adam Optimizer for supervision, as we are learning the distribution of prexels [5].

PressNet-Simple is trained with an initial learning rate of 10^{-4} for 40 epochs at a batch size of 128. PressNet-Simple takes 3 to 3.5 hours to train each LOSO data split on an NVIDIA TitanX GPU with 12GB of memory. The learning rate is reduced by 75% every 7 epochs, and MSE loss is used with the Adam Optimizer.

4 Evaluation and Visualization of Results

4.1 Quantitative Evaluation

KNN Baseline: KNN provides a convenient data-driven (memory-based, non-linear) way to directly map between two different modalities; in this case, pose to pressure. As the number of data samples becomes large, even simple NN (aka 1-NN) retrieval can perform surprisingly well, both theoretically [18] and empirically [65], due to the "Unreasonable Effectiveness of Data" [52]. The main drawback of KNN is the high cost of computing distances between a pose query and all samples in a large dataset; thus, we use it only as a baseline in our work.

The distance metric for KNN is the sum of Euclidean distances between corresponding normalized body joint locations. The foot pressure maps corresponding to these nearest neighbors are combined as a weighted average to generate the output pressure map prediction, using inverse distance weighting [60]. Empirical tests with K ranging from 1 to 50 showed that error reduces as K increases with diminishing improvements. Results from KNN-based pressure estimation, where K=5 and K=50, are included in Figure 7 for comparison with the deep learning based methods.

Table 4: Cross-view validation using PNS trained on camera View 1 to predict pressure maps from View 2 (Figure 2A). The reported metrics are Mean Absolute Error (MAE), Similarity (SIM), KL Divergence (KLD), and Information Gain (IG). Only outputs generated from input poses with 25 valid joints are included in the evaluation. Best values are shown in bold. Best is defined as closer to 1 for SIM and closer to 0 for MAE, KLD, and IG. Arrow indicates direction of better value.

Foot Pressure Error Metrics for Viewpoint Comparison (Mean / Std)								
Sub	MAE ↓		SIM ↑		KLD ↓		IG ↑	
#	View 1	View 2	View 1	View 2	View 1	View 2	View 1	View 2
1	**7.55** / **1.54**	9.15 / 1.94	**0.44** / **0.10**	0.36 / 0.12	**2.79** / **1.99**	4.27 / 3.23	**-0.74** / **0.50**	-1.00 / 0.74
2	**9.44** / 2.93	9.94 / **2.76**	0.23 / **0.15**	**0.25** / 0.16	5.21 / **5.38**	**4.65** / 5.72	-0.41 / 0.41	**-0.36** / **0.41**
3	8.82 / **2.37**	**8.82** / 2.51	0.39 / 0.13	**0.41** / **0.12**	2.95 / 2.70	**2.54** / **1.89**	-0.54 / 0.47	**-0.51** / **0.46**
4	**9.38** / **2.42**	9.76 / 2.55	**0.45** / **0.11**	0.45 / 0.12	**1.88** / **1.39**	2.16 / 1.53	**-0.52** / **0.45**	-0.57 / 0.51
5	**10.14** / **2.53**	10.15 / 2.61	0.47 / 0.13	**0.49** / **0.13**	1.94 / 1.52	**1.77** / **1.52**	-0.44 / 0.37	**-0.39** / **0.35**
6	**9.98** / **2.35**	10.67 / 2.45	**0.45** / **0.12**	0.44 / 0.12	1.54 / **1.35**	**1.41** / 1.44	-0.22 / 0.23	**-0.18** / **0.20**
7	10.18 / 2.23	**10.08** / **2.16**	0.34 / **0.09**	**0.36** / 0.11	4.44 / 2.61	**3.65** / **2.22**	-0.98 / 0.63	**-0.81** / **0.47**
8	8.31 / 3.58	**8.30** / **3.55**	0.33 / **0.13**	**0.36** / 0.14	3.32 / 2.63	**2.62** / **2.36**	-0.34 / 0.27	**-0.29** / **0.24**
9	8.24 / 2.88	**7.77** / **2.41**	0.37 / **0.13**	**0.40** / 0.13	2.82 / 2.58	**2.28** / **2.00**	-0.46 / 0.40	**-0.37** / **0.34**
10	**7.44** / **2.51**	7.68 / 2.53	0.44 / 0.14	**0.46** / **0.14**	3.18 / 3.22	**2.62** / **2.23**	-0.60 / 0.48	**-0.52** / **0.44**
Mean	**8.95** / **2.53**	9.23 / 2.55	0.39 / **0.12**	**0.40** / 0.13	3.01 / 2.54	**2.80** / **2.41**	-0.53 / 0.42	**-0.50** / **0.42**
Std	**0.98** / 0.50	1.00 / **0.40**	0.07 / 0.02	**0.07** / **0.01**	1.08 / **1.12**	**1.01** / 1.21	**0.20** / **0.11**	0.23 / 0.14

Evaluation Measures: We use six quantitative measures to evaluate the performance of the trained networks and KNN baseline:

1. Mean Absolute Error (MAE in kPa) between estimated foot pressure maps and measured ground truth pressure.
2. Three metrics for spatial distribution of the learned foot pressure map: Similarity, KL Divergence, and Information Gain [10].
3. Two measures on accuracy for estimated *Center of Pressure* (CoP) and *Base of Support* (BoS), which are directly related to the computation of stability (Figure 3C). We use ℓ_2 distance (in mm)/IoU (in %) between the estimated CoP from learned foot pressure maps and the CoP calculated directly from ground truth foot pressure to quantify CoP and BoS quality, respectively.

Evaluation of Predicted Pressure Maps: Table 3 shows our evaluation results using the first four metrics above on the KNN baseline and variations of PNS and PN. For each pressure prediction method, only frames that have 25 detected joints are included in the analysis to minimize confounding factors on the method's effectiveness. KL Divergence and Information Gain both show the advantages of networks on learning statistical distributions of the input data. The key takeaway is that both networks excel at predicting the spatial distribution of ground truth pressure more so than the overall magnitude.

Cross-view Validation. Our networks using 2D joint data were trained on View 1 (Figure 2A) of the two video cameras. Table 4 presents results of running a 2D network to predict foot pressure on images from the other camera, View 2. Results are similar to those in Table 3, indicating that both networks are

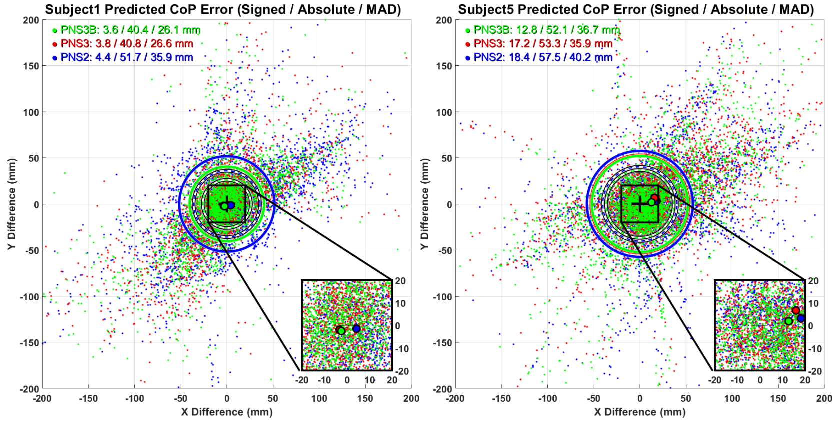
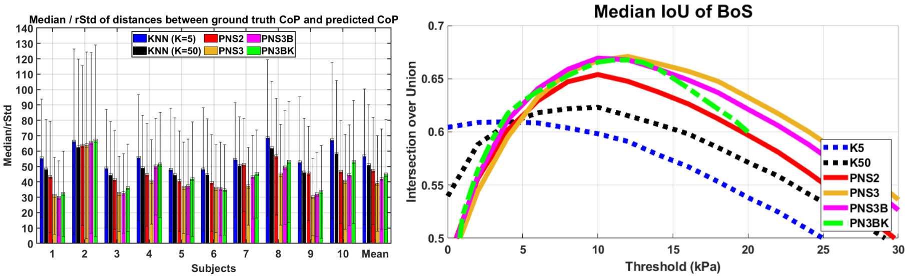


Fig. 6: 2D offsets between ground truth CoP (black cross) and CoP predicted by PNS2 (blue), PNS3 (red), and PNS3B (green). Large dots plot the mean of each scatter plot distribution; that they appear close to the Ground Truth (GT) indicates relatively symmetrical distributions of spatial error around the GT CoP. The concentric circles represent mean (solid) and median (dashed) offset distances as an error radius, and they indicate PNS3 and PNS3B CoP estimates cluster more tightly about the GT than PNS2.



(A) CoP

(B) BoS

Fig. 7: **(A)** Comparison of CoP offset distance errors across methods and subjects, characterized by robust estimation Median/rStd (robust STD). **(B)** Comparison of BoS using Intersection over Union (IoU) relative to ground truth over a range of pressure thresholds (Note: PN3BK has a normalized threshold scaled relative to kPa and therefore a shorter line). All results differ statistically significantly from one another. See more details in text.

robust to viewpoint and subject position/orientation relative to the camera view. PressNet-Simple appears less affected by viewpoint than PressNet, based on Similarity, KL Divergence, and Information Gain measures.

Distance to CoP: As a step towards estimating stability from pose, Center of Pressure (CoP) is computed from predicted foot pressure maps and compared to ground truth CoP locations computed from insole foot pressure readings. CoP is calculated as the weighted mean of the pressure elements (prexels) in the XY ground plane. A systematic threshold starting from 0 kPa is applied to both the ground truth and the predicted pressure, similar to the procedure used in [36]. Figure 6 provides scatter plots of CoP offset errors for each method. Also shown on the plot are the mean of the offset error distributions, and an error radius for each method derived from the mean and median of the offset distances. This figure highlights the similarities across subjects as well as the clear improvement that both deep learning networks make over KNN in more accurately predicting CoP.

Figure 7A presents a robust analysis of outcomes where central location X,Y is estimated by 2D geometric median, computed by Weiszfeld's algorithm [69]. The spread of data is estimated by a robust standard deviation measure (rStd), derived as median absolute deviation (MAD) from the median, multiplied by a constant 1.4826 that scales MAD to be a *consistent estimator* of population standard deviation [58]. Bar height in the chart corresponds to median CoP offset distances, and the whiskers on each median bar represent rStd. Median and rStd values, which are generally smaller than mean and std because robust estimators suppress the effects of outliers. We computed p-values between all pairs of methods, finding that all outcomes differ statistically significantly ($p << 0.001$), except PNS3/PNS3B ($p{=}0.009$), even with large variances, which makes sense since our train/test set sizes are very large (Figure 2B). However, the conclusion about relative merits of each method remain the same, with both proposed networks outperforming the KNN baseline. It should be noted that Subject 2 consistently under-performs for all methods, which may be an indication of inaccuracy in the input pose or ground truth pressure data, requiring further investigation. Figure 7B presents an analysis of the Base of Support (BoS) resulting from the predicted foot pressures relative to the ground truth pressure using the Intersection over Union (IoU). With identical overlap (IoU = 1) being the goal, the results indicate that all networks outperform KNN (K5 and K50) with the PNS2 under-performing all 3D methods with 65-68% overlap. The X-axis presents the threshold (in kPa) used to calculate the BoS for comparison for all but the PN3BK model, which is on a normalized and unitless scale as part of data processing for KL Divergence loss. PN3BK is scaled relative in Figure 7B to provide easy visual comparison.

4.2 Qualitative Evaluation

Figure 8 visualizes ground truth, foot pressure predictions, and their BoS and CoP for some sample frames. For each subject, the foot pressure predictions and ground truth are rendered with independent pressure scales (weight related) based on the pressure range needed for each subject. In addition to the qualitative comparison by visualization, the respective mean absolute errors with respect to ground truth frames have been calculated and included in Table 3.

Finally, we show preliminary results on exploring the potential use of estimated foot-pressure distributions to obtain classic stability measures defined in

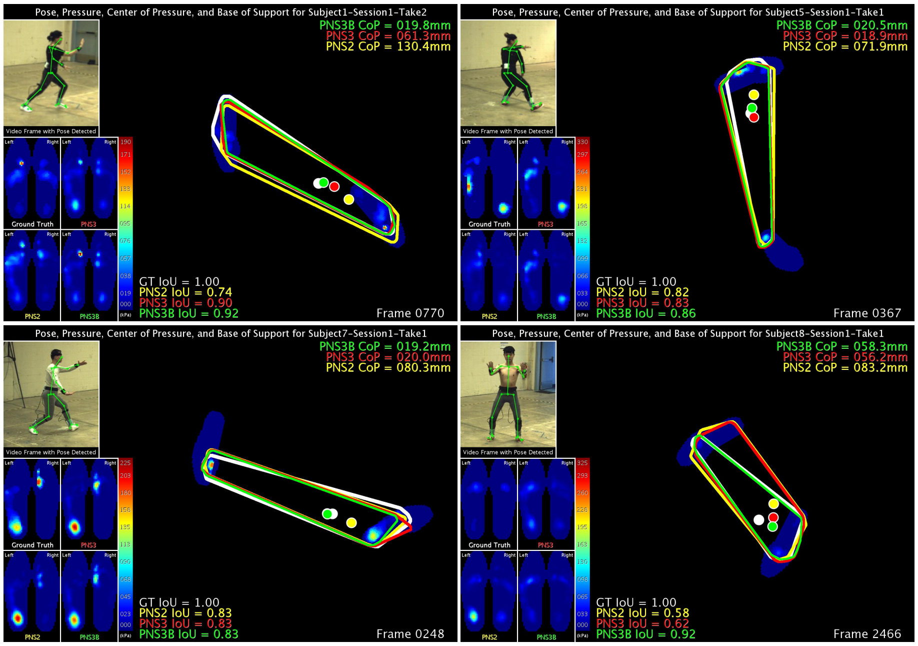
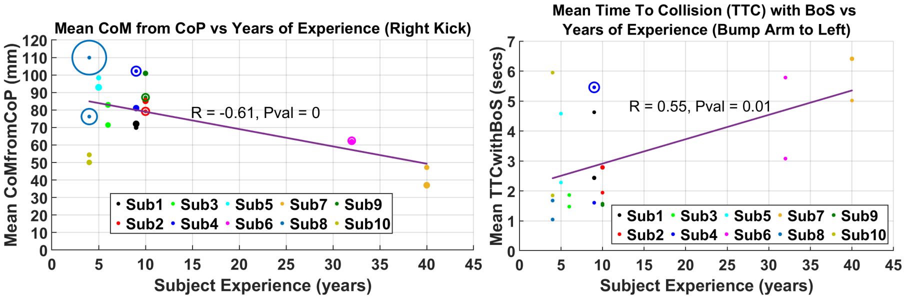


Fig. 8: Sample output frames showing the ground truth and estimated Center of Pressure (CoP) and Base of Support (BoS). Foot pressure is scaled for each subject based on their range of pressure. BoS and CoP of Ground Truth (white), PNS (yellow), PNS3 (red), and PNS3B (green) plotted as an overlaid on the floor plane. Intersection over Union (IoU) and distance to Ground Truth CoP (mm) are used to quantify the quality of BoS and CoP estimation, respectively.



(A) CoM from CoP: R = -0.61

(B) TTC with BoS: R = 0.55

Fig. 9: Correlation between the number of years of Taiji experience/training N and two different stability metrics can be seen here on two different Taiji poses. They seem to confirm the general observation that the more experienced Taiji practitioners are more stable, where N is larger: **(A)** distance between CoM and CoP is smaller; and **(B)** the time to reach the boundary of BoS is longer.

kinesiology [13,26–28,35,41]. Motivated by findings in medicine via randomized trials that Taiji intervention may improve stability of certain populations [21], we observe convincing correlations between years of Taiji practice and two stability measures (Figure 3C) for two different Taiji poses (Figure 9). The trend of correlation is consistent with previous work; as with [21], the more Taiji practice, the more stable a subject is. Figure 9A shows negative correlation, meaning CoM and CoP align better for more experienced Taiji subjects. Figure 9B shows positive correlation, meaning: the most experienced subjects can maintain better stability when CoM's Time To Collision (TTC) with BoS is longer.

5 Conclusions

We present a fully validated approach to estimate foot pressure distributions from 2D/3D human body pose. Given the multi-faceted complexity of this pose-to-foot pressure mapping problem, we have gained several insights from this exercise: (1) KNN is a reasonable baseline predictor, while deep learning networks surpass KNN statistically significantly; (2) 3D pose input has a high positive impact over 2D, albeit with an upfront higher computational cost; (3) system performance is surprisingly stable across subject weight, gender, height variations, and the number of subjects in the training/testing data; (4) networks trained on one camera view produce comparable results when tested on images from a different view, confirming that the Taiji dataset provides adequate orientation generalization information; (5) quantitative evaluations on a subset of *ordinary poses* indicates that networks trained on Taiji movements can be generalized to non-Taiji-specific poses; and (6) correlation between the quantified results of deep learning networks support our initial hypothesis that learning a mapping from kinematics to dynamics from static images is feasible, opening up a door for precision computer vision devoted to human body centered sciences. Access to implementation and dataset details are available through the project website: http://vision.cse.psu.edu/research/dynamicsFromKinematics/index.shtml.

6 Acknowledgments

We thank our collaborator Professor John Challis for offering his Kinesiology perspectives and helpful discussions. We thank Master Sitan Chen of Sitan Tai Chi School; Dr. Hesheng Bao of Win-Win Kung Fu Culture Center, Inc.; Professor Zuofeng Sun of School of Physical Education, Hebei Normal University; and Professor Pingping Xie of Tianjin University of Sport for their expertise and support of many years. We thank all the human subject volunteers for participating in our Taiji data collection. This human study is carried out under Penn State University IRB #8085 and was supported in part by NSF grant IIS-1218729 and the College of Engineering Dean's office of Penn State University.

References

1. Agarwal, A., Triggs, B.: 3D human pose from silhouettes by relevance vector regression. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). vol. 2, pp. 882–888 (2004) 3
2. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2D human pose estimation: New benchmark and state of the art analysis. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). p. 3686–3693 (2014) 3, 4
3. Arvin, M., Hoozemans, M., Pijnappels, M., Duysens, J., Verschueren, S., Van Dieen, J.: Where to step? Contributions of stance leg muscle spindle afference to planning of mediolateral foot placement for balance control in young and older adults. Frontiers in Physiology **9**, 1134 (2018) 3
4. Bächer, M., Whiting, E., Bickel, B., Sorkine-Hornung, O.: Spin-it: Optimizing moment of inertia for spinnable objects. ACM Trans. Graph. **33**(4), 96:1–10 (2014) 3
5. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg (2006) 9
6. Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., Black, M.J.: Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) European Conference on Computer Vision (ECCV). LNCS, vol. 9905, pp. 561–578. Springer (2016) 3
7. Brubaker, M.A., Sigal, L., Fleet, D.J.: Estimating contact dynamics. In: IEEE International Conference on Computer Vision (ICCV). pp. 2389–2396 (2009) 3
8. Brubaker, M., Fleet, D., Hertzmann, A.: Physics-based person tracking using the anthropomorphic walker. Int J Comput Vis (IJCV) **87**(1), 140–155 (2010) 3
9. Bulat, A., Tzimiropoulos, G.: Human pose estimation via convolutional part heatmap regression. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) European Conference on Computer Vision (ECCV). LNCS, vol. 9905, pp. 717–732. Springer (2016) 1, 3
10. Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., Durand, F.: What do different evaluation metrics tell us about saliency models? IEEE Trans Pattern Analysis and Machine Intelligence (PAMI) **41**(3), 740–757 (March 2019) 10
11. Cai, Y., Ge, L., Cai, J., Yuan, J.: Weakly-supervised 3D hand pose estimation from monocular RGB images. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) European Conference on Computer Vision (ECCV). LNCS, vol. 11210, pp. 678–694. Springer, Cham (2018) 3
12. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1302–1310 (2017) 1, 3, 5
13. Chaudhry, H., Bukiet, B., Ji, Z., Findley, T.: Measurement of balance in computer posturography: Comparison of methods - a brief review. Journal of bodywork and movement therapies **15**(1), 82–91 (2011) 14
14. Chen, C.H., Ramanan, D.: 3D human pose estimation= 2D pose estimation + matching. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 5759–5767 (2017) 3
15. Chen, W., Wang, H., Li, Y., Su, H., Wang, Z., Tu, C., Lischinski, D., Cohen-Or, D., Chen, B.: Synthesizing training images for boosting human 3D pose estimation. In: IEEE Intl. Conf. on 3D Vision (3DV). pp. 479–488 (2016) 1, 3
16. Chen, X., Yuille, A.L.: Articulated pose estimation by a graphical model with image dependent pairwise relations. In: Advances in Neural Information Processing Systems (NIPS). pp. 1736–1744 (2014) 1, 3

17. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2017) 8
18. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory **13**(1), 21–27 (1967) 9
19. Eckardt, N., Rosenblatt, N.J.: Healthy aging does not impair lower extremity motor flexibility while walking across an uneven surface. Human Movement Science **62**, 67–80 (2018) 3
20. Fan, X., Zheng, K., Lin, Y., Wang, S.: Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1347–1355 (2015) 1, 3
21. Fuzhong, L., Peter, H., Kathleen, F., Elizabeth, E., Ronald, S., Johnny, G., Gianni, M., Sara, S.: Tai Chi and postural stability in patients with Parkinson's disease. New England Journal of Medicine **366**(6), 511–519 (2012) 14
22. Gilbert, A., Trumble, M., Malleson, C., Hilton, A., Collomosse, J.: Fusing visual and inertial sensors with semantics for 3D human pose estimation. Int J Comput Vis (IJCV) **127**, 381—397 (2019) 3
23. Grimm, R., Sukkau, J., Hornegger, J., Greiner, G.: Automatic patient pose estimation using pressure sensing mattresses. In: Handels, H., Ehrhardt, J., Deserno, T., Meinzer, H., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin, pp. 409–413. Springer, Berlin, Heidelberg (2011) 3
24. Güler, R.A., Neverova, N., Kokkinos, I.: Densepose: Dense human pose estimation in the wild. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 7297–7306 (2018) 1, 3
25. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems (NIPS). pp. 5767–5777 (2017) 7
26. Hof, A., Gazendam, M., Sinke, W.: The condition for dynamic stability. Journal of biomechanics **38**(1), 1–8 (2005) 14
27. Hof, A.L.: The equations of motion for a standing human reveal three mechanisms for balance. Journal of Biomechanics **40**(2), 451–457 (2007) 3, 4, 14
28. Hof, A.L.: The "extrapolated center of mass" concept suggests a simple control of balance in walking. Human Movement Science **27**(1), 112–125 (2008) 3, 4, 14
29. Hsiao, H., Guan, J., Weatherly, M.: Accuracy and precision of two in-shoe pressure measurement systems. Ergonomics **45**(8), 537–555 (2002) 6
30. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2261–2269 (2017) 3
31. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proc. Intl. Conf. on Machine Learning (ICML). vol. 37, pp. 448–456 (2015) 8
32. Ionescu, C., Li, F., Sminchisescu, C.: Latent structured models for human pose estimation. In: IEEE International Conference on Computer Vision (ICCV). pp. 2220–2227 (2011) 3, 4
33. Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3.6m: Large scale datasets and predictive methods for 3D human sensing in natural environments. IEEE Trans Pattern Analysis and Machine Intelligence (PAMI) **36**(7), 1325–1339 (jul 2014) 3, 4
34. Iqbal, U., Milan, A., Gall, J.: Posetrack: Joint multi-person pose estimation and tracking. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2011–2020 (2017) 3, 4

35. Jian, Y., Winter, D.A., Ishac, M.G., Gilchrist, L.: Trajectory of the body COG and COP during initiation and termination of gait. Gait & Posture **1**(1), 9–22 (1993) 14
36. Keijsers, N., Stolwijk, N., Nienhuis, B., Duysens, J.: A new method to normalize plantar pressure measurements for foot size and foot progression angle. Journal of Biomechanics **42**(1), 87–90 (2009) 12
37. Ko, J.H., Wang, Z., Challis, J.H., Newell, K.M.: Compensatory mechanisms of balance to the scaling of arm-swing frequency. Journal of Biomechanics **48**(14), 3825–3829 (2015) 3
38. Lemaire, E.D., Biswas, A., Kofman, J.: Plantar pressure parameters for dynamic gait stability analysis. In: IEEE Engineering in Medicine and Biology Society (EMBS). pp. 4465–4468 (2006) 3
39. Li, Z., Sedlar, J., Carpentier, J., Laptev, I., Mansard, N., Sivic, J.: Estimating 3D motion and forces of person-object interactions from monocular video. In: IEEE Conf. Computer Vision and Pattern Recognition (CVPR). pp. 8632–8641 (2019) 3
40. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) European Conference on Computer Vision (ECCV). LNCS, vol. 8693, p. 740–755. Springer, Cham (2014) 3, 4
41. Lugade, V., Lin, V., Chou, L.S.: Center of mass and base of support interaction during gait. Gait & Posture **33**(3), 406–411 (2011) 14
42. Lv, X., Chai, J., Xia, S.: Data driven inverse dynamics for human motion. ACM Trans. Graph. **35**(6), 1–12 (2016) 3
43. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: ICML Workshop on Deep Learning for Audio, Speech and Language Processing (2013) 8
44. Martinez, J., Hossain, R., Romero, J., Little, J.J.: A simple yet effective baseline for 3D human pose estimation. In: IEEE International Conference on Computer Vision (ICCV). pp. 2659–2668 (2017) 3, 8
45. McKay, M.J., Baldwin, J.N., Ferreira, P., Simic, M., Vanicek, N., Wojciechowski, E., Mudge, A., Burns, J.: Spatiotemporal and plantar pressure patterns of 1000 healthy individuals aged 3–101 years. Gait & Posture **58**, 78–87 (2017) 3
46. Moreno-Noguer, F.: 3D human pose estimation from a single image via distance matrix regression. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1561–1570 (2017) 3
47. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) European Conference on Computer Vision (ECCV). LNCS, vol. 9912, pp. 483–499. Springer (2016) 1, 3
48. Nie, B.X., Wei, P., Zhu, S.C.: Monocular 3D human pose estimation by predicting depth on joints. In: IEEE International Conference on Computer Vision (ICCV). pp. 3467–3475 (2017) 3
49. Pai, Y.C.: Movement termination and stability in standing. Exercise and Sport Sciences Reviews **31**(1), 19–25 (2003) 3, 4
50. Pataky, T., Mu, T., Bosch, K., Rosenbaum, D., Goulermas, J.: Gait recognition: Highly unique dynamic plantar pressure patterns among 104 individuals. Journal of The Royal Society Interface **9**, 790–800 (2012) 3
51. Pavlakos, G., Zhou, X., Derpanis, K.G., Daniilidis, K.: Coarse-to-fine volumetric prediction for single-image 3D human pose. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1263–1272. IEEE (2017) 3
52. Pereira, F., Norvig, P., Halevy, A.: The unreasonable effectiveness of data. IEEE Intelligent Systems **24**(02), 8–12 (mar 2009) 9

53. Prévost, R., Bächer, M., Jarosz, W., Sorkine-Hornung, O.: Balancing 3D models with movable masses. In: Conf. on Vision, Modeling and Visualization (VMV'16). pp. 9–16. Eurographics Association (2016) 3
54. Prévost, R., Whiting, E., Lefebvre, S., Sorkine-Hornung, O.: Make it stand: Balancing shapes for 3D fabrication. ACM Trans. Graph. **32**(4), 81:1–10 (2013) 3
55. Putti, A., Arnold, G., Abboud, R.: Foot pressure differences in men and women. Foot and Ankle Surgery **16**(1), 21–24 (2010) 3
56. Ravichandran, B.: BioPose-3D and PressNet-KL: A Path to Understanding Human Pose Stability from Video. Master's thesis, Computer Science and Engineering, The Pennsylvania State University (2020) 6
57. Rogez, G., Weinzaepfel, P., Schmid, C.: LCR-Net++: Multi-person 2D and 3D pose detection in natural images. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) **42**(5), 1146–1161 (2020) 3
58. Rousseeuw, P.J., Croux, C.: Alternatives to the median absolute deviation. Journal of the American Statistical Association **88**(424), 1273–1283 (1993) 12
59. Seethapathi, N., Wang, S., Saluja, R., Blohm, G., Körding, K.P.: Movement science needs different pose tracking algorithms. CoRR **abs/1907.10226** (2019) 1, 3
60. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proc. ACM National Conference (ACM'68). pp. 517–524 (1968) 9
61. Sigal, L., Balan, A., Black, M.J.: HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International Journal of Computer Vision **87**(1), 4–27 (Mar 2010) 3, 4
62. Simo-Serra, E., Ramisa, A., Alenyà, G., Torras, C., Moreno-Noguer, F.: Single image 3D human pose estimation from noisy observations. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2673–2680 (2012) 3
63. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 5693–5703 (2019) 6
64. Sun, X., Shang, J., Liang, S., Wei, Y.: Compositional human pose regression. In: IEEE Intl. Conf. on Computer Vision (ICCV). pp. 2621–2630 (2017) 3
65. Tatarchenko, M., Richter, S.R., Ranftl, R., Li, Z., Koltun, V., Brox, T.: What do single-view 3D reconstruction networks learn? In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 3405–3414 (2019) 9
66. Tompson, J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: Advances in Neural Information Processing Systems (NIPS). pp. 1799–1807 (2014) 3
67. Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C.: Efficient object localization using convolutional networks. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 648–656 (2015) 8
68. Toshev, A., Szegedy, C.: Deeppose: Human pose estimation via deep neural networks. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1653–1660 (2014) 1, 3
69. Vardi, Y., Zhang, C.H.: The multivariate L_1-median and associated data depth. Proceedings of the National Academy of Science **97**(4), 1423–1426 (2000) 12
70. Vera-Rodriguez, R., JSD., M., Fierrez, J., Ortega-Garcia, J.: Comparative analysis and fusion of spatiotemporal information for footstep recognition. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) **35**, 823–34 (2013) 3
71. Vondrak, M., Sigal, L., Jenkins, O.C.: Physical simulation for probabilistic motion tracking. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1–8 (2008) 3

72. Wang, C., Bannuru, R., Ramel, J., Kupelnick, B., Scott, T., Schmid, C.: Tai Chi on psychological well-being: Systematic review and meta-analysis. BMC complementary and alternative medicine **10**, 23 (2010) 4, 5
73. Winter, D.A.: Human balance and posture control during standing and walking. Gait & Posture **3**, 193–214 (1995) 1
74. Zhou, X., Zhu, M., Leonardos, S., Derpanis, K.G., Daniilidis, K.: Sparseness meets deepness: 3D human pose estimation from monocular video. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 4966–4975 (2016) 3
75. Zhou, X., Huang, Q., Sun, X., Xue, X., Wei, Y.: Towards 3D human pose estimation in the wild: a weakly-supervised approach. In: IEEE International Conference on Computer Vision (ICCV). pp. 398–407 (2017) 3