

STUTTGART MEDIA UNIVERSITY

MASTER THESIS

**Applied Research of an End-to-End
Human Keypoint Detection Network with
Figure Ice Skating as Application Scope**

Author:

Nadin-Katrin APEL

Supervisor:

Prof. Dr. J. MAUCHER

Prof. Dr. S. RADICKE

*A thesis submitted in fulfillment of the requirements
for the degree*

Master of Science

Computer Science and Media

May 18, 2020

Declaration of Authorship

I, Nadin-Katrin APEL, declare that this thesis titled, “Applied Research of an End-to-End Human Keypoint Detection Network with Figure Ice Skating as Application Scope” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at
- this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such
- quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was
- done by others and what I have contributed myself.

Signed:

Date:

“Data is a precious thing and will last longer than the systems themselves.”

Tim Berners-Lee

STUTTGART MEDIA UNIVERSITY

Abstract

Computer Science and Media

Master of Science

Applied Research of an End-to-End Human Keypoint Detection Network with Figure Ice Skating as Application Scope

by Nadin-Katrin APEL

Human joint detection is a key component for machines to understand human actions and behaviors. Especially in figure ice skating this understanding is an indispensability, where there are many difficult figures and poses, even difficult to clearly understand for the professionalized jury. Herewith we present an end-to-end approach to detect the 2D poses of a person in images and videos. In the architecture we combine three branches: Image Segmentation, Body Part Detection, and Human joint detection. The applied research reveals multiple findings which outperform current existing main players with the special application scope of figure ice skating.

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor

Contents

Declaration of Authorship	iii
Abstract	vii
Acknowledgements	ix
Acronyms	xiii
1 Introduction	1
1.1 Motivation and Goals	2
1.2 Related Work	3
2 Figure Skating Pose Detection	7
2.1 Complexity of Figures	7
2.2 Distinct Rating System	8
3 Dataset	9
3.1 Figure Skating Dataset	9
3.1.1 XSens: Inertial motion capturing recordings on the ice rink . . .	10
3.2 Synthetic Dataset: 3DPeople	12
3.3 Data Processing	13
4 Method	15
4.1 Network Architecture	15
4.1.1 Body Part Detection Module	15
4.1.2 Joint Detection Module	15
4.2 Training Performance	16
4.3 Inference Runtime Analysis	16
4.4 Implementation Details	16
5 Experiments	19
5.1 Ablation Study	20
5.1.1 Body Parts Module	20
5.1.2 Joint Module	20
5.2 Comparison of Optimizer Algorithms	20
5.3 Performance of loss functions	22
5.3.1 Sparse Categorical Cross Entropy	23
5.3.2 Mean Squared Error	23
5.3.3 Our custom loss function CILoss	23
6 Conclusion and future thoughts	25
Bibliography	27

Acronyms

bl layer L with largest feature maps. [20](#)

List of Figures

5.1	Predicted Mask	19
5.2	HPNet v7	20
5.3	Accuracy	21
5.4	Correct BPR	21
5.5	loss	21
5.6	Learning Rate SGD	22
5.7	Loss SGD	22
5.8	Accuracy SGD	22
5.9	CILoss	23

List of Tables

List of Abbreviations

LAH List Abbreviations **Here**
WSF What (it) Stands For

Physical Constants

use it Speed of Light $c_0 = 2.997\,924\,58 \times 10^8 \text{ m s}^{-1}$ (exact)

List of Symbols

a	distance	m
P	power	W (J s ⁻¹)
ω	angular frequency	rad

For/Dedicated to/To my...

Chapter 1

Introduction

Human 2D pose estimation has gained more and more attraction in recent years. For example Facebook, one of the BigFive technology companies, has published 73 research paper targeting the problem of pose estimation in the last three years. The most popular ones are DensePose, VideoPose3d or Mask R-CNN [2, 12, 15, 26]. A company in Canada *wrnch.ai* even specialized on keypoint recognition from image and video data with a lot of product options [42]. Furthermore, many enterprises are becoming more and more interested in Sport Content Analysis (SPA) e.g. Bloomberg, SAP and Panasonic, just naming a few [11, 31].

But how got this topic into such a demanded focal point? Probably this is due to the various application areas in which Pose estimation can be encountered. Main fields are sports, visual surveillance, autonomous driving, entertainment, health care and robotics [25, 34, 47]. For example Vaak, a japanese startup, developed a software, which would detect shoplifters, even before they were able to remove items from a store. This yielded in a drastic reduction of stealing crimes in stores.

The exercise of sport not via visiting a sports course, gym or club became of fundamental severance in 2020, when the Coronavirus SARS-CoV-2 spread the entire world [30]. Many courses such as Yoga, Pilates or general fitness routines went online and were often conducted via Zoom, Instagram live or other video streaming technologies [3]. However, what participants were often missing, was the feedback of the coach on how the exercise was going, and whether it was done right or wrong. So in 2020 more than ever was missed a technology which is good at pose estimation, or even further, action recognition, in sports.

Most investigations in this field target usual activities not including complex poses which can be encountered in professional sport. This is why these architectures

often fail when applied to more complex movements. For competitive sports there are various metrics of high interest depending on the environment. Competition and training can be differentiated as can be sports executed by multiple athletes versus single combats. Basketball or soccer as team sports for example are interested on predictions about how the other team behaves during the game and which would be the best reaction to their behaviour for winning the game. During practice 2d pose recognition can help to optimize the sports-person movements by taking the role of a coach. This could provide an answer to the question on how certain activities might be optimized? Single competitive sports with very complex movement routines are for example gymnastics and figure ice skating. Both sports include various artistic body movements, which are not part of daily activities. Even famous and well rated 2d pose recognition networks such as OpenPose or VideoPose3d fail to recognize these poses.

If this problem was solved it could help with action recognition and support during practice or relieve the jury on competitions. A predictor could for example suggest, what an athlete should do to land a certain jump. On the other hand jury is rare and the job sitting all day in the ice-rink on weekends with only a very small salary is not very attractive. Furthermore, people often complain scoring is not executed fairly.

Especially in figure ice skating an accurate 2d pose recognition could make a huge contribution. This is why this paper investigates 2d pose recognition with special focus on figure ice skating.

Don't forget to sum thesis up here

1.1 Motivation and Goals

A working 2d pose estimator could make a huge contribution to figure ice skating. Especially when building an action recognizer on top of it. However, as of today, this was not possible yet due to the complex poses and the different gliding movements on the ice. Especially spins with their fast rotation and stretching poses are of high complexity to these estimators. Such an estimator could support fair scoring during competitions or help to improve motion sequences during practice.

With the downward trend of jury staff and the increasing demand for more small competitions, jury is asked more and more in figure ice skating. Particularly the role of the technical specialist or controller diagnosing the individual elements on the ice is of high demand. Some competitions were even canceled in recent years, because they were not able to find the according jury. Furthermore, sitting all day in the cold ice rink for only a very low salary is not attractive at all. These long demanding days challenge concentration and many competition participants often complain about jury not rating fairly enough, completely forgetting the demanding work the jury has to do. Here a 2d pose estimator could contribute by recognizing the different elements or even scoring. This would not only relieve the jury but also could increase fairness.

During practice 2d pose estimators could examine the specific motions during elements and give hints how to improve these. Probably they could even suggest certain exercises to learn an element like a spin, jump or certain step. Additionally they could keep track of training and provide analysis metrics to the skaters and coaches.

All in all 2d pose estimation is very interesting not only because of all the possible different appliance possibilities in this sport, but as well because of the challenging task to build an according estimator, which was not possible until today.

dolor

1.2 Related Work

2D Pose estimation sets the baseline for machines to understand actions. It is the problem of localizing human joints or keypoints in images and videos. Many research studies explored and researched this topic already with the most popular ones being OpenPose and VideoPose3d [5, 26].

For 2D pose recognition there are mainly two general procedures: either top-down or bottom-up. Top-down first detects a person and then finds their keypoints. Whereas bottom-up first detects all keypoints in the image and then refers the corresponding people. For top-down it is argued, that if a person is not detected via a bounding box or alike, no keypoints can be found. This would lead to more

undetected keypoints in the frames of a video. When there are many people in the image with many occlusions the people often can not be detected. However, when a person is correctly detected, it is said that accuracy would be higher [24].

A famous top-down approach for example is Mask R-CNN developed by the Facebook AI Research team. It consists of three branches and two stages. The first stage presents the Region Proposal Network (RPN). It proposes candidate bounding boxes for objects. The second stage performs classification and bounding box regression by extracting features using region of interest pooling, which they refer as RoiPool. Additionally, Mask R-CNN predicts a binary mask for each ROI in the second stage. They receive top results in the COCO challenges for instance segmentation, bounding box object detection, and person keypoint detection [12]. Other famous top-down approaches include *Simple Baselines*, *the Cascaded Pyramid Network* or *Deep High-Resolution Learning* [8, 35, 43].

One of the most discussed and popular bottom-up approaches as of today is OpenPose. Their net predicts vector fields for the joint connections, which they call part affinity fields. Additionally, it estimates candidate keypoint locations via gaussian distributions. These they refer as part confidence maps. From these detections they refer the associate human poses. OpenPose shows very good results on the MPII and COCO challenges. They highlight their performance, which especially shows it's strength when detecting multiple people. The performance wouldn't change even if more and more people enter the scene. With the according hardware this would even show decent results in realtime [5]. In a newer research, they additionally pay attention to the temporal characteristic in video sequences in their work of Spatio Temporal Fields. Their approach is able to track multiple people's poses across frames being runtime-invariant to the number of people in the frames. Furthermore, they receive highly competitive results on the PoseTrack challenges [28]. Some other famous bottom-up approaches include *Convolutional Pose Machines* and *PifPaf* [16, 41].

The COCO, MPII and PoseTrack challenges lead to several studies in the pose estimation field. However, their dataset targets rather simple daily activities. There have been only conducted a few studies on competitive sports such as basketball, ice-hockey or swimming [10, 24, 40]. For sports including full body flexibility or special unconventional jump or turn rotations as can be found in dance, gymnastics

or figure ice skating, there have been only a few studies [6, 17, 45, 46]. These studies encounter three main problems in the figure skating domain: The first is domain knowledge. In C. Xu et Al. research they try to predict the technical and performance scores from video data with the only the ice skating program as video input and the judge scores as labels [45]. This will very likely not result in useful results since first, the figure skating judging system adjusts every year, second, there are always different judges on the competitions who all have their own rating style, and third however the skater is one of the always winning ones, this skater can do a good program falling at the main elements and still score very well, because the jury knows this skater and is just human in their judgement. This is one of the main controversies in the figure ice skating fairness of program judgement. Another problem is the missing dataset. There didn't exist a dataset with joint labeling until FSD-10 [17], which only came out at the beginning of the year. One very interesting study from Yu, Ri et Al. tries to create simulations from the figure ice skating elements. They exactly encountered the problem, that pose estimation currently does not work on difficult pose sequences such as spins or very flexible positions. However, they were able to successfully predict jumps and simple steps from videos into 3d simulation [43].

Our Work

In our work we investigated the performance of VideoPose3d, OpenPose and wrnch.ai on figure ice skating elements. We elaborated the creation of a figure ice skating dataset with the help of the XSens motion capturing data in a real figure ice skating arena environment. Furthermore, we evaluated the creation of a dataset from these motion capturing data with the help of Blender and Makehuman. Furthermore, we created an end-to-end fully convolutional architecture consisting of three modules for background extraction, body part and keypoint detection. Since our work concentrates on pose estimation in single skating, the background extraction module is an indispensability. Multiple experiments with the network architectures, learning optimizers and loss functions resulted in decent results, which can be applied in figure ice-skating and run on usual hardware. At the end we elaborated several future thoughts were these studies can continue to.

Chapter 2

Figure Skating Pose Detection

2.1 Complexity of Figures

Figure ice skating includes very special movement sequences which stand out from other sports. Thanks to the surface of the ice, skaters perform gliding movements. Furthermore, their programs include highly artistic movements with explosive take-offs, very high rotation speeds and flexible positions. Current pose estimators such as OpenPose, VideoPose3d or wrnch.ai all fail to correctly predict spins with their high rotation speed and difficult flexible positions. In this section we try to explain why especially figure ice skating means a challenge in human pose estimation.

In single figure ice skating, there are three main element types in a competitive program, which receive technical scores: jumps, steps, and spins. There are seven different listed jumps, which only differ in their take-off phase. Three of these are jumped just from one edge, the Axel, Loop and Salchow. The other ones additionally use the skate toe as a catapult. Furthermore, they can be distinguished by the edge that is last skated on the ice before the skater takes off. This is one of the reasons, why many skaters have problems with the Lutz and Flip jumps. For the jury it is often hard to tell as well, whether the jump should get an edge deduction. All the jumps can include a different amount of rotation in the air, whereas four rotations was the maximum a skater could perform until today. All these jumps can be combined in various manners, include features such as lifted arms or difficult steps before or after the element to increase the level of difficulty, resulting in higher scores.

For spins there are four main positions: upright, sit, camel and layback. These positions as well can be combined with various features, as for example jumps or

difficult elastic positions such as the famous Biellmann spin.

There are plenty of different step sequence elements including turns on the ice and steps. These as well can be combined with multiple additional features resulting in higher scores.

The above named elements only explain the absolute basics in figure ice skating. In practice scoring and the creation of ice skating programs is much more difficult. Nevertheless, this illustrates nicely, that however for the not professionalized audience many elements look the same, there are plenty of different metrics and elements. Moreover, each skater has their own style and performs these elements slightly different. This is what makes it so hard in machine-learning to correctly predict the elements. Which is why it is important that the basis, the keypoint recognition module, has to predict the poses correctly. Because otherwise the action recognition module has no chance to make correct estimations.

2.2 Distinct Rating System

- isu scores [37] - isu guidelines/ level of execution goe [36] - human struggle as well
-> rating system with points, many abstractions, still often experienced as not fair

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Chapter 3

Dataset

3.1 Figure Skating Dataset

The correct dataset with the according labels is what decides weather a neural network architecture will be able to sucessfully learn keypoints from image data or video sequence data. As of today, there does not exist a publicly available dataset which labels keypoints to ice skating performances or elements. Famous existing one such as FIS-V or the MIT datasets only include the total technical and performance scores of the complete performances, which are not applicable her as explained in [chapter 2](#).

To create a reasonable dataset for figure ice skating there are various possibilities. Most straightforward would be to make use of the huge amout of vidos published everyday on online platforms such as YouTube and label the videos by hand. This however would cost a huge amount of time and includes outliers coming from human errors.

Another possibility would be to work with motion capture. There are three main methods for recording motion capturing data. Markerless capture is a method where videos are recorded via a single or multiple depth cameras. This comes with the advantage that the athletes are not distracted or feel uncomfortable by any markers [13, 18, 29]. When working with markers there are two sensinging types: optical or inertial. Optical markers are reflective markers, which are attached to characteristic part of the body. These reflections are tracked via multiple cameras to make sure the motion movements are taken especially when a person is moving, so the markers are not concealed [39]. Inertial motion capture uses IMU sensors that are attached to body parts [23, 44]. 3D positions then are calculated via multiple sensor metrics. XSens system for example includes several trackers which contain Magnetometers,

gyroscopes and accelerometers. Via their MVN Animate and Anlaze systems the tracked data is postprocessed on a biomechanical model and then can be used in game engines such as Unity or Unreal, or in 3D animation software such as Blender or 3DSMax [1].

In an environment such as the ice rink motion capture has to deal with several difficulties. For one, the lighting is often not smooth and there are several different light sources. Moreover, it is very difficult to get the whole ice rink for a recording, because most of the time the ice rink is fully booked with figure ice skating, ice-hockey, public skating or short track speed skating. Further, many ice rinks are governed by the local community [9], or are highly expensive to rent. This makes markerless and optical recordings very challenging. On the other hand inertial recordings as by the XSens technology seem very promising. Even when there are multiple skaters on the ice, recordings can be tracked in usual practice time slots.

3.1.1 XSens: Inertial motion capturing recordings on the ice rink

In our research study we tested recordings on the ice rink with the MVN Link product from XSens. This product includes 17 wired IMU motion trackers which can be either attached into a suit or onto straps. The sensors are connected via cables to a battery with a life time of 9.5 h. These trackers are then connected via a router to their MVN Animate software. The wireless range of the trackers is 50 to 150 meters, however the trackers are able to buffer the recordings and transmit these later when there is a connection to the hub. Their MVN Animate software helps with calibration and shows the recording results as soon as the trackers are close enough for transmission. Furthermore, it processes the data and performs adjustments in a postprocessing step [22].

Our experiences with MVN Link from XSens

The overall setup was very time consuming. To correctly attach all the straps and align the cables took us a minimum of 20 minutes. Here we tried out the suit and the straps with equal preparation times. Then the calibration on the ice with the MVN Analyze software was very time consuming again and took about 20 minutes. The calibration process had to be repeated several times, when the calibration failed. These problems probably came from the different gliding movement on the ice. So the skaters had to mimic usual floor movements for the calibration to work correctly.

Since the recordings are very time consuming we could not work with young professional competitive skaters, because the setup was too drawn-out. We then took recordings from a senior competitive skater. The internal method was very beneficial here, because other skaters on the ice did not intervene in the recordings. Additionally, however sometimes the connection was lost to the hub, when the skater came back the recordings were transmitted correctly. The postprocessing from their software did work really good as well. What was disruptive however, were the sensors and battery, which the skater had to carry. This made the movements feel constrained. Further, in figure ice skating falls are common. A fall on the battery or one of the sensors would have been harmful, which again influenced the movement of the skaters. In comparison there are other products on the market, that promise a faster setup e.g. through a suit in which the sensors are integrated without cables [23]. These products would very likely better fit for figure ice skating motion captures.

Create dataset with Blender and MakeHuman

We processed the recorded data from the XSens trackers in MVN Animate and then exported a short sequence as bvh file to test possible dataset creation procedures. Further, we used the open source MakeHuman [20] to create a figure skating avatar. Here we used the figure skating dress and ice skates, which are freely available from the MakeHuman community [38]. With the help of the MakeHuman plugin in Blender 2.7 [4] we then imported the created avatar from MakeHuman and retargeted the bvh motion capture file onto the rig. We then set up a default scene with lighting and a moving camera. To obtain the keypoints, we wrote a Python script in Blender to calculate the joint locations in relation to the camera view and exported the resulting keypoints into numpy files.

However, the above described process seemed to be smooth to conduct, we did face a couple of challenges. For one, artistic elastic movements yielded in errors on the armature, since the joints were restricted to certain regions of movement, which seemed to work for usual daily activities but not figure ice skating ones. Another one was that with the moving camera the keypoints were not calculated correctly and drawn with a small offset. Recent research studies suggest to export the animation as bvh files and later calculate the joint positions with a program such as Matlab [7, 27, 32]. So this could improve the calculation of joint positions.

Our conclusions on creating a synthetic dataset in figure ice skating

We are convinced that the above described process to create a synthetic dataset for figure ice skating is an indispensability when it comes to figure ice skating, because of the following reasons: first: with inertial motion tracking is very accurate especially in an environment such as an ice rink when recording can be easily done during practice sessions. second: much less recordings are necessary thanks to the large arbitrary degrees of variation which is possible with an animation software such as blender third: the diversity of data is exceptionally vast. From one small motion capture file multiple different avatars can be retargetted, female and male, age variations, clothes and race variances. Further, multiple different light sources and camera views can be applied and the background randomly selected. fourth: calculated labels by the according animation software or post-processing program are just more exact than human labeled data.

All these points confirm the legitimacy of synthetic datasets in figure ice skating. Which is why we researched on already existing synthetic datasets since in the scope of this thesis creating an own dataset was not possible due to the huge time efforts. The synthetic dataset 3DPeople as described in the next section correspond to our expectations of a synthetic dataset.

3.2 Synthetic Dataset: 3DPeople

The dataset 3DPeople was created via created motion capture files. They took 70 realistic action sequences from Mixamo [21]. The sequences contain on average 110 frames and range from usual activities with little motions such as driving or drinking until break dance moves or backflips. These actions they then retarget onto 80 armatures which are created with MakeHuman [20] or Adobe Fuse [19]. There are 40 female and 40 male characters with plenty of variation in body shapes, skin tones, outfits and hair. Furthermore, they record the actions with a projective camera and 800 mm focal length from four different viewpoints which are orthogonally aligned with the ground. The distance to the subjects vary arbitrary as do light sources and static background images. In sum the dataset includes 22,400 clips with the rendered video sequences as 640x480 pixel image frames, the according depth maps, optical flow and semantic information such as body parts and cloth labels.

We found this dataset highly sophisticated due to the wide variance and diversity of the data. Especially in terms of clothes it stands out from other famous datasets

such as Human3.6M [33]. They use a variety of wide and tight clothes. For example do they include dresses and shorts. There variance is greater compared to their freely available combats [14, 33].

3.3 Data Processing

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Chapter 4

Method

4.1 Network Architecture

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

4.1.1 Body Part Detection Module

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

4.1.2 Joint Detection Module

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

4.2 Training Performance

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

4.3 Inference Runtime Analysis

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

4.4 Implementation Details

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in.

Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Chapter 5

Experiments

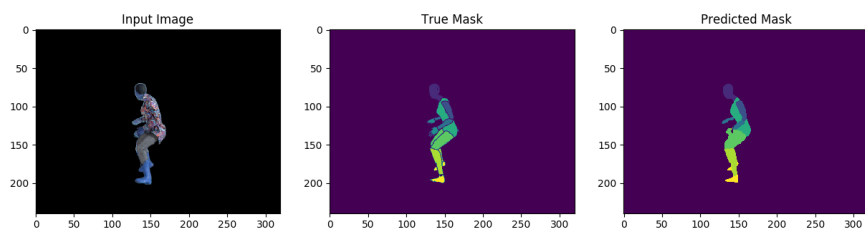


FIGURE 5.1: Predicted mask after 3845th epoch with custom loss function and Adam optimizer_kps

5.1 Ablation Study

5.1.1 Body Parts Module

Stride-down, -up convolution before [bl](#)

MobileNet extended with UNet

MobileNet extended with HPNet

Experiment with concat and add layers

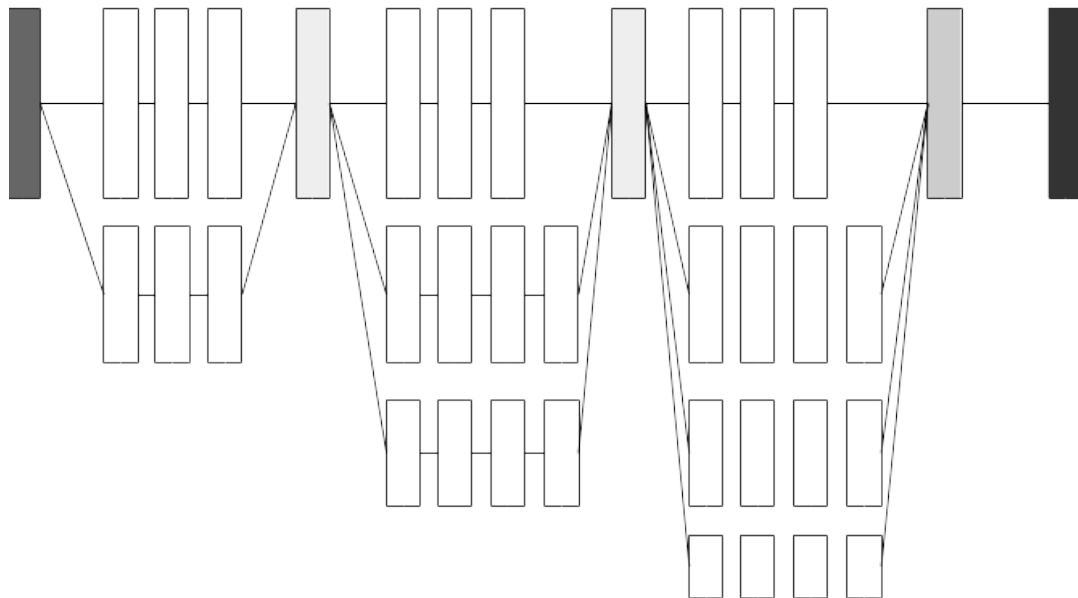


FIGURE 5.2: HPNet v7.

Best performing network HPNet v7

5.1.2 Joint Module

Dense Modules

Fully Convolutional

5.2 Comparison of Optimizer Algorithms

- Adam

- Nadam
- SGD

constant learning rate

Constant decreasing learning rate

Constant decreasing learning rate with reset of learning rate on plateau

Increasing decreasing learning rate on plateau

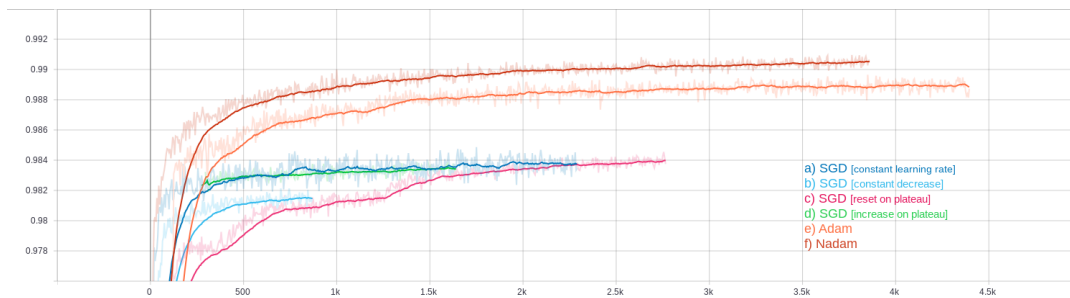


FIGURE 5.3: Accuracy

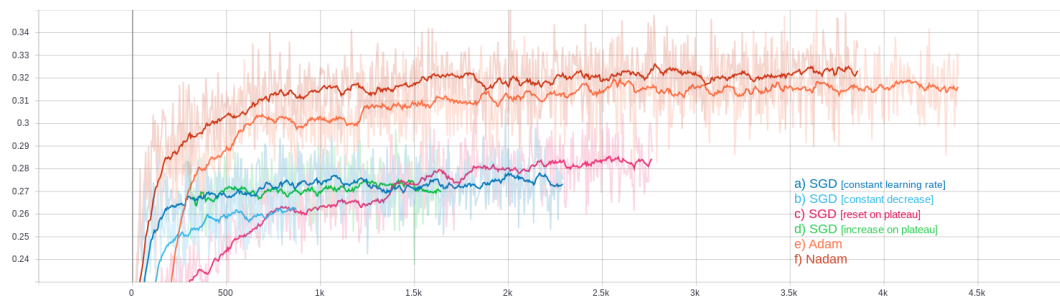


FIGURE 5.4: Correct body part pixel relation



FIGURE 5.5: Loss

Comparison of Adam, Nadam and SGD

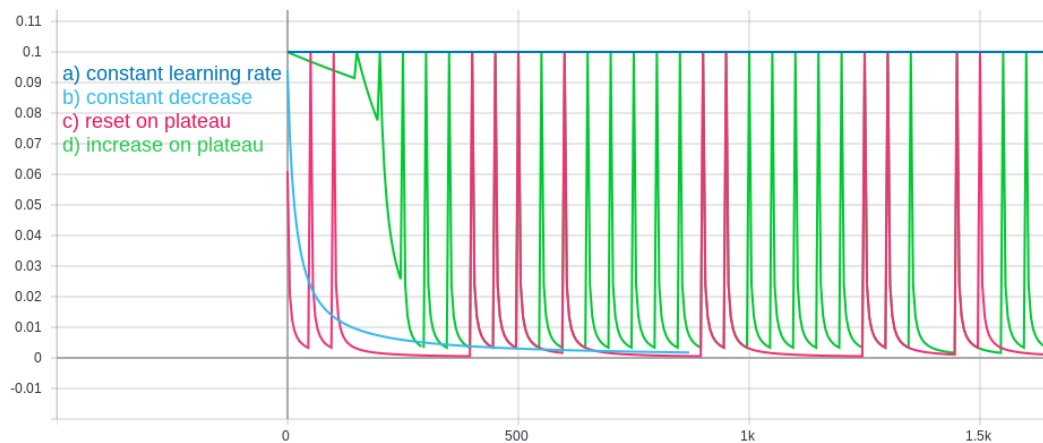


FIGURE 5.6: Learning Rate SGD.

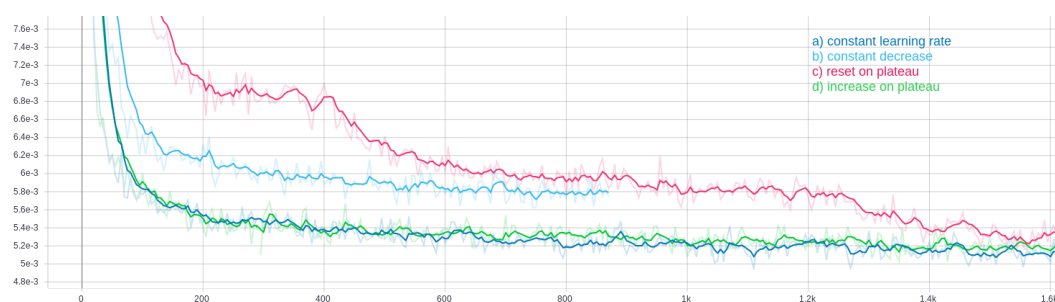


FIGURE 5.7: Loss SGD.

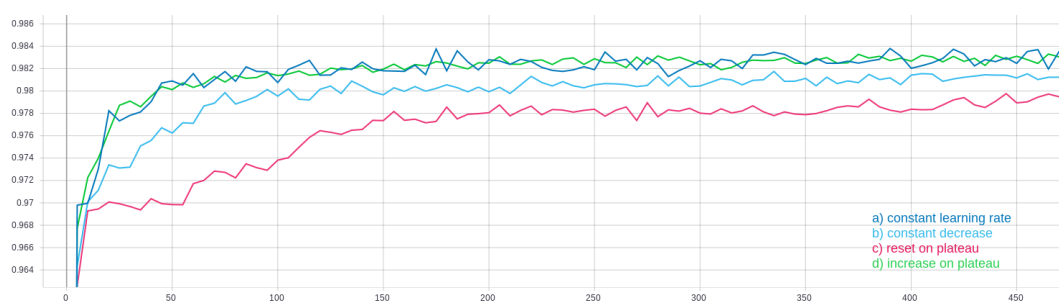


FIGURE 5.8: Accuracy SGD.

Experiments with SGD

5.3 Performance of loss functions

All performance measures are conducted on the Nadam optimizer_kps with the HP-Net for body part recognition from Recognition of body parts [5.1.1](#)

5.3.1 Sparse Categorical Cross Entropy

5.3.2 Mean Squared Error

5.3.3 Our custom loss function CILoss

This loss function confronts the problem of class imbalance, which especially occurs in body part recognition. The background pixels appear most often, and the different body part classes occur by far less often and even they differentiate a lot in their relative occurrence.

We try to confront this problem with a weighed map, which takes the body parts as a graph and calculates the distances from each body part b_x to all other body parts b_n , and stores this data inside a table.

Additionally this weight map is evened out with a multiplier to reduce the distances and facilitate the learning process for the network.

$$\theta = y_t(x) - y_p(x)$$

$$\delta = \theta * \mu[\operatorname{argmax}(y_t)]$$

$$L = \sum_{i=0}^n \theta_i + \delta_i$$

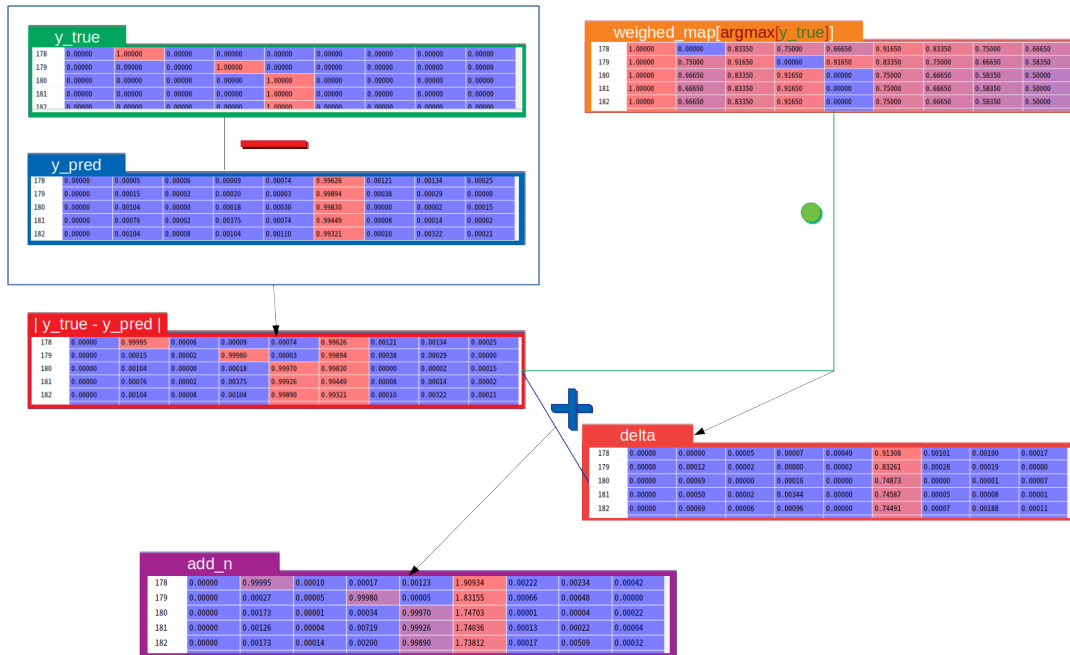


FIGURE 5.9: Visualization of custom loss calculation

Chapter 6

Conclusion and future thoughts

Bibliography

- [1] 2017. URL: <https://base.xsens.com/hc/en-us/community/posts/115003027934-How-do-the-Xsens-suits-actually-work> (visited on 05/17/2020).
- [2] Iasonas Kokkinos Rĩ za Alp Güler Natalia Neverova. “DensePose: Dense Human Pose Estimation In The Wild”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).
- [3] Sabrina Barr. “CORONAVIRUS: FROM YOGA TO BARRY’S BOOTCAMP — BEST EXERCISE CLASSES ON ZOOM, INSTAGRAM AND YOUTUBE”. In: *Independent* (2020). URL: <https://www.independent.co.uk/life-style/health-and-families/coronavirus-home-workout-exercise-class-yoga-dance-kids-elderly-joe-wicks-a9421126.html> (visited on 05/11/2020).
- [4] *Blender.today - daily art & development live streams*. URL: <https://www.blender.org/> (visited on 05/17/2020).
- [5] Z. Cao et al. “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019), pp. 1–1.
- [6] Daniel Castro et al. “Let’s Dance: Learning From Online Dance Videos”. In: (2018).
- [7] Christoforos Charalambous and Anil Bharath. “A data augmentation methodology for training machine/deep learning gait recognition algorithms”. In: (Oct. 2016).
- [8] Y. Chen et al. “Cascaded Pyramid Network for Multi-person Pose Estimation”. In: (2018), pp. 7103–7112.
- [9] *Eiswelt Stuttgart - Stadt Stuttgart*. URL: <https://www.stuttgart.de/eiswelt> (visited on 05/17/2020).
- [10] M. Fani et al. “Hockey Action Recognition via Integrated Stacked Hourglass Network”. In: (2017), pp. 85–93.

- [11] David Fox. "Video-based sports analytics system from SAP and Panasonic announced at IBC". In: *SVG europe* (2014). URL: <https://www.svg-europe.org/blog/headlines/video-based-sports-analytics-from-sap-and-panasonic-announced-at-ibc/> (visited on 05/11/2020).
- [12] K. He et al. "Mask R-CNN". In: (2017), pp. 2980–2988.
- [13] <https://thecaptury.com/>. *THE CAPTURE - Markerless Motion Capture*. 2020. URL: <https://thecaptury.com/> (visited on 05/17/2020).
- [14] *Human3.6M*. URL: <http://vision.imar.ro/human3.6m/description.php> (visited on 05/17/2020).
- [15] Facebook Artificial Intelligence. *Facebook publications with topic pose estimation*. 2020. URL: [https://ai.facebook.com/results/?q=pose%20estimation&content_types\[0\]=publication&years\[0\]=2020&years\[1\]=2019&years\[2\]=2018&sort_by=relevance&view=list&page=1](https://ai.facebook.com/results/?q=pose%20estimation&content_types[0]=publication&years[0]=2020&years[1]=2019&years[2]=2018&sort_by=relevance&view=list&page=1) (visited on 05/03/2020).
- [16] S. Kreiss, L. Bertoni, and A. Alahi. "PifPaf: Composite Fields for Human Pose Estimation". In: (2019), pp. 11969–11978.
- [17] Shenlan Liu et al. "FSD-10: A Dataset for Competitive Sports Content Analysis". In: (2020).
- [18] Z. Liu et al. "Template Deformation-Based 3-D Reconstruction of Full Human Body Scans From Low-Cost Depth Cameras". In: *IEEE Transactions on Cybernetics* 47.3 (2017), pp. 695–708.
- [19] *Make custom 3D characters for your Photoshop projects*. URL: <https://www.adobe.com/products/fuse.html> (visited on 05/17/2020).
- [20] *MakeHuman - Open Source tool for making 3D characters*. URL: <http://www.makehumancommunity.org/> (visited on 05/17/2020).
- [21] *Motion Pack*. URL: <https://www.mixamo.com/#/?page=1&type=Motion%2CMotionPac> (visited on 05/17/2020).
- [22] *MVN Animate*. URL: <https://www.xsens.com/products/mvn-animate> (visited on 05/17/2020).
- [23] Perception Neuron. *Perception Neuron Motion Capture*. URL: <https://neuronmocap.com/> (visited on 05/17/2020).

- [24] Takuya Ohashi, Yosuke Ikegami, and Yoshihiko Nakamura. "Synergetic Reconstruction from 2D Pose and 3D Motion for Wide-Space Multi-Person Video Motion Capture in the Wild". In: (2020).
- [25] Paritosh Parmar and Brendan Tran Morris. "Learning to Score Olympic Events". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2017), pp. 76–84.
- [26] D. Pavllo et al. "3D Human Pose Estimation in Video With Temporal Convolutions and Semi-Supervised Training". In: (2019), pp. 7745–7754.
- [27] Albert Pumarola et al. "3DPeople: Modeling the Geometry of Dressed Humans". In: (Apr. 2019).
- [28] Y. Raaj et al. "Efficient Online Multi-Person 2D Pose Tracking With Recurrent Spatio-Temporal Affinity Fields". In: (2019), pp. 4615–4623.
- [29] Radical. *The Body in Motion - No suits. No hardware.* 2020. URL: <https://getrad.co/> (visited on 05/17/2020).
- [30] RKI. "SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19)". In: *Robert Koch Institut* (2020). URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html (visited on 05/11/2020).
- [31] Bloomberg Press Room. "Bloomberg Sports Launches "Stats Insights," Sports Analysis Blog". In: (2012). URL: <https://www.bloomberg.com/company/press/bloomberg-sports-launches-stats-insights-sports-analysis-blog/> (visited on 05/11/2020).
- [32] Nikolaos Sarafianos et al. "3D Human Pose Estimation: A Review of the Literature and Analysis of Covariates". In: *Computer Vision and Image Understanding* 152 (Sept. 2016). DOI: [10.1016/j.cviu.2016.09.002](https://doi.org/10.1016/j.cviu.2016.09.002).
- [33] Leonid Sigal, Alexandru Balan, and Michael Black. "HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion". In: *International Journal of Computer Vision* 87 (Mar. 2010), pp. 4–27. DOI: [10.1007/s11263-009-0273-6](https://doi.org/10.1007/s11263-009-0273-6).
- [34] Waqas Sultani, Chen Chen, and Mubarak Shah. "Real-World Anomaly Detection in Surveillance Videos". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 6479–6488.

- [35] K. Sun et al. "Deep High-Resolution Representation Learning for Human Pose Estimation". In: (2019), pp. 5686–5696.
- [36] International Skating Union. "Levels of Difficulty and Guidelines for marking Grade of Execution and Program Components, Season 2020/21". In: *Communication No. 2324: SINGLE & PAIR SKATING* (2020). URL: <https://www.isu.org/inside-isu/isu-communications/communications/24332-2324-sp-levels-of-difficulty-and-guidelines-for-marking-goe-final/file>.
- [37] International Skating Union. "Scale of Values, Levels of Difficulty and Guidelines for marking Grade of Execution, season 2018/19". In: *Communication No. 2168: SINGLE & PAIR SKATING* (2018). URL: <https://www.isu.org/docman-documents-links/isu-files/documents-communications/isu-communications/17142-isu-communication-2168/file>.
- [38] *User contributed assets*. URL: http://www.makehumancommunity.org/content/user_contributed_assets.html (visited on 05/17/2020).
- [39] Vicon. *Award Winning Motion Capture Systems*. 2020. (Visited on 05/17/2020).
- [40] B. Victor et al. "Continuous Video to Simple Signals for Swimming Stroke Detection with Convolutional Neural Networks". In: (2017), pp. 122–131.
- [41] S. Wei et al. "Convolutional Pose Machines". In: (2016), pp. 4724–4732.
- [42] "wrnchAI, BUILT FOR ALL YOUR HUMAN VISION NEEDS". In: (2020). URL: <https://wrnch.ai/product> (visited on 05/12/2020).
- [43] B. Xiao, H. Wu, and Y. Wei. "Simple Baselines for Human Pose Estimation and Tracking". In: (2018).
- [44] XSens. *Xsens 3D motion tracking*. URL: <https://www.xsens.com/> (visited on 05/17/2020).
- [45] C. Xu et al. "Learning to Score Figure Skating Sport Videos". In: *IEEE Transactions on Circuits and Systems for Video Technology* (2019), pp. 1–1.
- [46] Ri Yu, Hwangpil Park, and J. Lee. "Figure Skating Simulation from Video". In: *Computer Graphics Forum* 38 (Oct. 2019), pp. 225–234. DOI: [10.1111/cgf.13831](https://doi.org/10.1111/cgf.13831).
- [47] Nan Zhao et al. "See your mental state from your walk: Recognizing anxiety and depression through Kinect-recorded gait data". In: *PLoS ONE* 14 (2019).