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 13, 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

Acronyms

bl layer **L** with largest feature maps. 14

List of Figures

List of Tables

List of Abbreviations

LAH	**L**ist **A**bbreviations **H**ere
WSF	**W**hat (it) **S**tands **F**or

Physical Constants

use it Speed of Light	$c_0 = 2.997\,924\,58 \times 10^{8}\ \mathrm{m\,s^{-1}}$ (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 and VideoPose3d [1, 5, 8]. Furthermore, many enterprises are becoming more and more interested in Sport Content Analysis (SPA) e.g. Bloomberg, SAP and Panasonic, just naming a few [4, 10].

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 [7, 11, 13]. 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 [9]. 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 [2]. 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.

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 [3, 8]. A popular company in Canada *wrnch.ai* even specialized on keypoint recognition from image and video data with a lot of product options [12].

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 unlabeled frames in 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 [6]. OpenPose, as the famous bottom-up approach, shows results where they were able to detect multiple people with their poses in videos. With the according hardware this would even show decent results in realtime [3].

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.

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 event 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 in 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.

1.2 Related Work

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.

Chapter 2

Figure Skating Pose Detection

2.1 Complexity of Figures

- existing KP detectors struggle (OpenPose, VideoPose)
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.

2.2 Distinct Rating System

- 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

Existing Datasets - missing annotation - reference paper

Create Figure Skating Dataset with Motion Capture and Blender - XSens

3.2 Synthetic Dataset: 3DPeople

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.

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

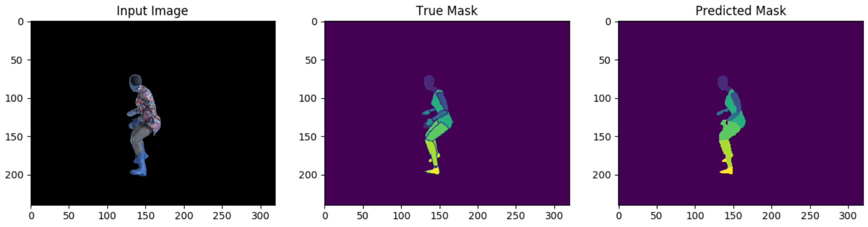


FIGURE 5.1: Predicted mask after 3845th epoch with custom loss function and Adam optimizer_kps

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

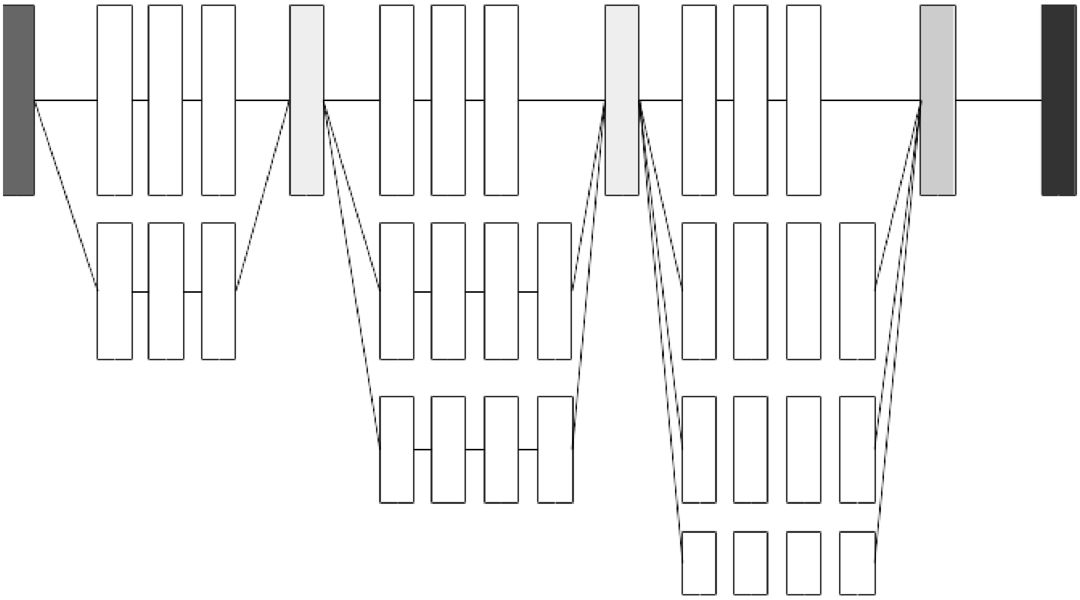


FIGURE 5.2: HPNet v7.

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

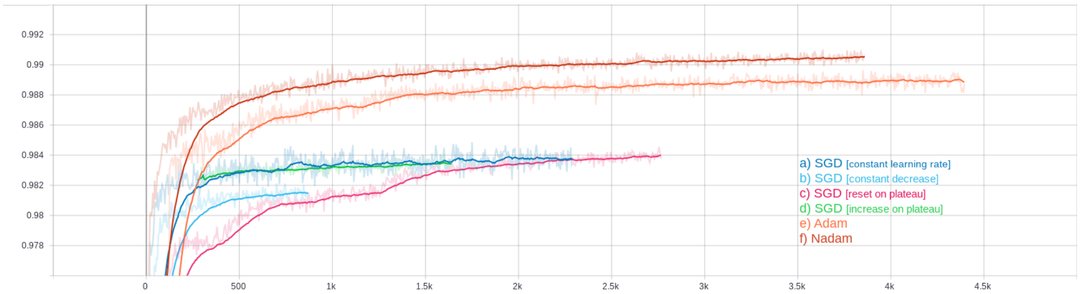


FIGURE 5.3: Accuracy

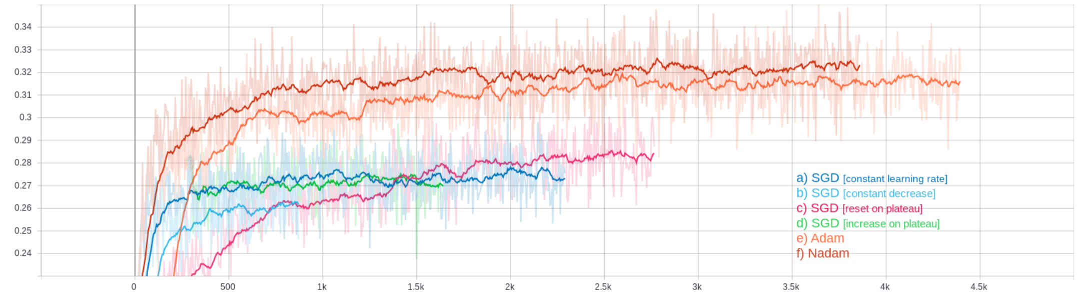


FIGURE 5.4: Correct body part pixel relation

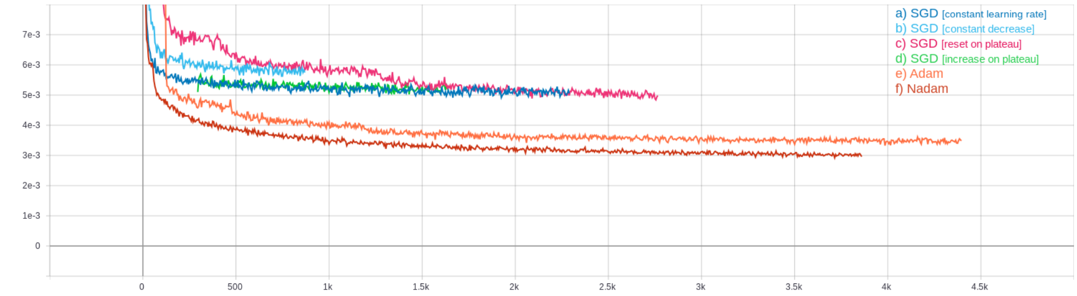


FIGURE 5.5: Loss

Comparison of Adam, Nadam and SGD

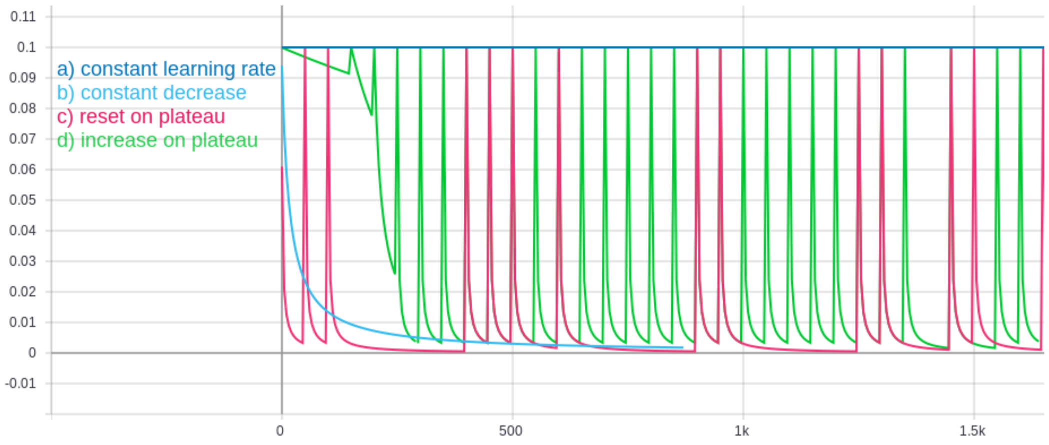


FIGURE 5.6: Learning Rate SGD.

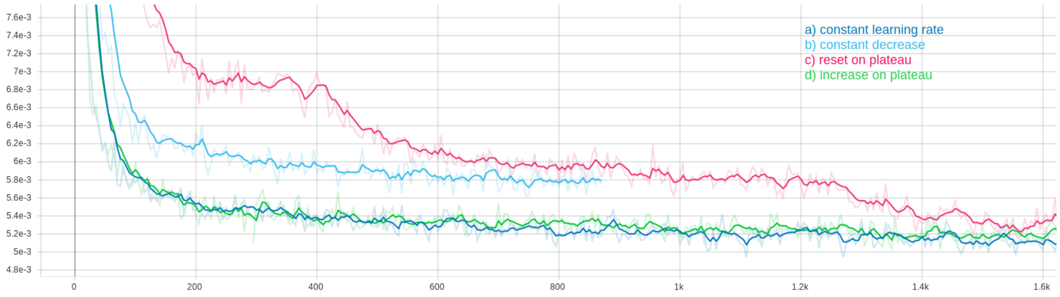


FIGURE 5.7: Loss SGD.

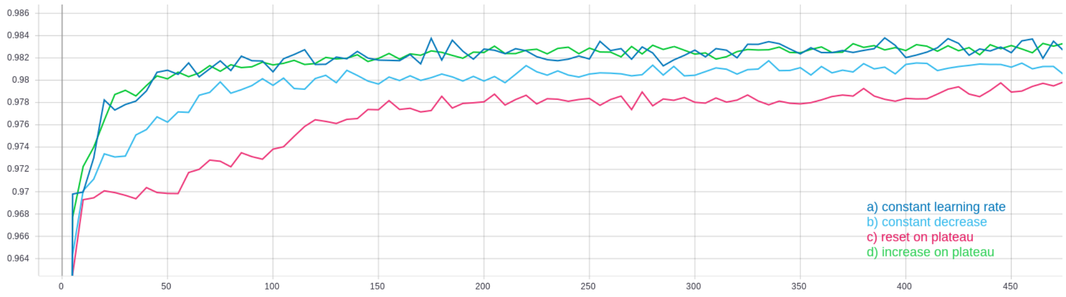


FIGURE 5.8: Accuracy SGD.

Experiments with SGD

5.3 Performance of loss functions

All performance measures are conducted on the Nadam optimizer_kps with the HPNet 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 event they differentiate a lot in their relative occurrence.

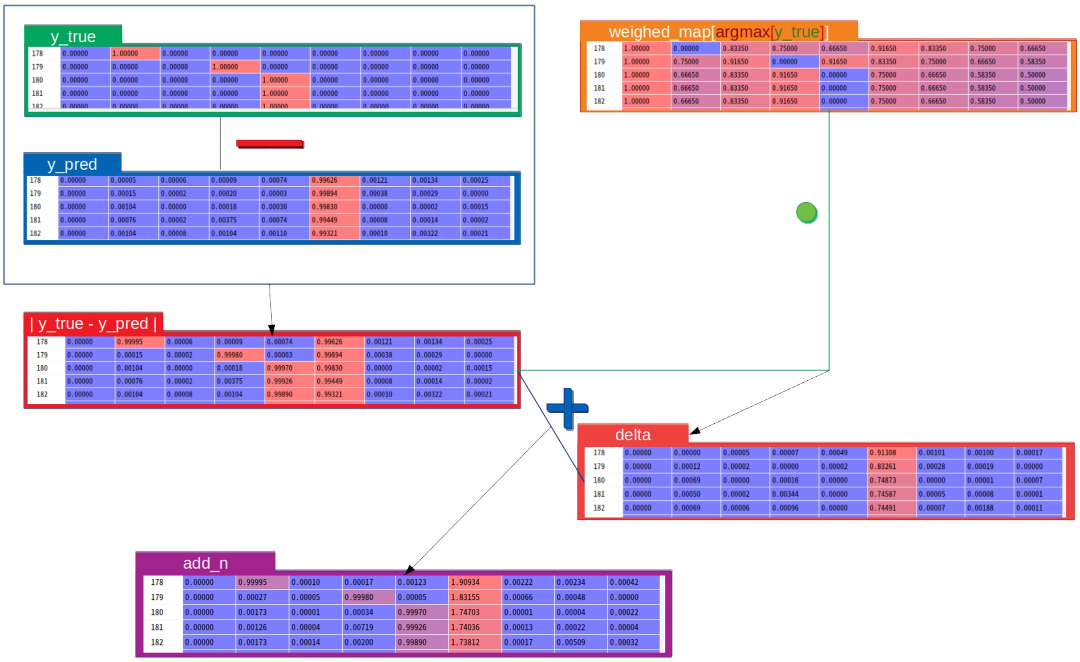
We try to confront this problem with a weighed map, which takes the body parts as a graph and calculates the distances from each body part b_x to all other body parts b_n, and stores this data inside a table.

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

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

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

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



y_true

178	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
179	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000
180	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000
181	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000
182	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000

y_pred

178	0.00000	0.00005	0.00006	0.00009	0.00074	0.99626	0.00121	0.00134	0.00025
179	0.00000	0.00015	0.00002	0.00020	0.00003	0.99894	0.00038	0.00029	0.00000
180	0.00000	0.00104	0.00000	0.00018	0.00030	0.99830	0.00000	0.00002	0.00015
181	0.00000	0.00076	0.00002	0.00375	0.00074	0.99449	0.00008	0.00014	0.00002
182	0.00000	0.00104	0.00008	0.00104	0.00110	0.99321	0.00010	0.00322	0.00021

weighed_map[argmax[y_true]]

178	1.00000	0.00000	0.83350	0.75000	0.66650	0.91650	0.83350	0.75000	0.66650
179	1.00000	0.75000	0.91650	0.00000	0.91650	0.83350	0.75000	0.66650	0.58350
180	1.00000	0.66650	0.83350	0.91650	0.00000	0.75000	0.66650	0.58350	0.50000
181	1.00000	0.66650	0.83350	0.91650	0.00000	0.75000	0.66650	0.58350	0.50000
182	1.00000	0.66650	0.83350	0.91650	0.00000	0.75000	0.66650	0.58350	0.50000

| y_true - y_pred |

178	0.00000	0.99995	0.00006	0.00009	0.00074	0.99626	0.00121	0.00134	0.00025
179	0.00000	0.00015	0.00002	0.99980	0.00003	0.99894	0.00038	0.00029	0.00000
180	0.00000	0.00104	0.00000	0.00018	0.99970	0.99830	0.00000	0.00002	0.00015
181	0.00000	0.00076	0.00002	0.00375	0.99926	0.99449	0.00008	0.00014	0.00002
182	0.00000	0.00104	0.00008	0.00104	0.99890	0.99321	0.00010	0.00322	0.00021

delta

178	0.00000	0.00000	0.00005	0.00007	0.00049	0.91308	0.00101	0.00100	0.00017
179	0.00000	0.00012	0.00002	0.00000	0.00002	0.83261	0.00028	0.00019	0.00000
180	0.00000	0.00069	0.00000	0.00016	0.00000	0.74873	0.00000	0.00001	0.00007
181	0.00000	0.00050	0.00002	0.00344	0.00000	0.74587	0.00005	0.00008	0.00001
182	0.00000	0.00069	0.00006	0.00096	0.00000	0.74491	0.00007	0.00188	0.00011

add_n

178	0.00000	0.99995	0.00010	0.00017	0.00123	1.90934	0.00222	0.00234	0.00042
179	0.00000	0.00027	0.00005	0.99980	0.00005	1.83155	0.00066	0.00048	0.00000
180	0.00000	0.00173	0.00001	0.00034	0.99970	1.74703	0.00001	0.00004	0.00022
181	0.00000	0.00126	0.00004	0.00719	0.99926	1.74036	0.00013	0.00022	0.00004
182	0.00000	0.00173	0.00014	0.00200	0.99890	1.73812	0.00017	0.00509	0.00032

FIGURE 5.9: Visualization of custom loss calculation

Chapter 6

Conclusion and future thoughts

Bibliography

[1] Iasonas Kokkinos Rı̃za Alp Güler Natalia Neverova. "DensePose: Dense Human Pose Estimation In The Wild". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

[2] Sabrina Barr. "CORONAVIRUS: FROM YOGA TO BARRY'S BOOTCAMP — BEST EXERCISE CLASSES ON ZOOM, INSTAGRAM AND YOUTUBE". In: *Indewpendent* (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).

[3] 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.

[4] David Fox. "Video-based sports analytics system from SAP and Panasonic announced at IBC". In: *SVG europe* (2014). URL: https://www.svgeurope.org/blog/headlines/video-based-sports-analytics-from-sap-and-panasonic-announced-at-ibc/ (visited on 05/11/2020).

[5] Facebook Artificial Intelligence. "Feacebook publications with topic pose estimation". In: (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 (visited on 05/03/2020).

[6] 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).

[7] 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.

[8] D. Pavllo et al. “3D Human Pose Estimation in Video With Temporal Convolutions and Semi-Supervised Training”. In: (2019), pp. 7745–7754.

[9] 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).

[10] 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).

[11] 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.

[12] “wrnchAI, BUILT FOR ALL YOUR HUMAN VISION NEEDS”. In: (2020). URL: https://wrnch.ai/product (visited on 05/12/2020).

[13] 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).