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 2, 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 1
1.2 Related Work 1

2 Figure Skating Pose Detection **3**
2.1 Complexity of Figures 3
2.2 Distinct Rating System 3

3 Dataset **5**
3.1 Synthetic Dataset: 3DPeople 5
3.2 Data Processing 5

4 Method **7**
4.1 Network Architecture 7
4.1.1 Body Part Detection Module 7
4.1.2 Joint Detection Module 7
4.2 Training Performance 8
4.3 Inference Runtime Analysis 8
4.4 Implementation Details 8

5 Experiments **11**
5.1 Ablation Study 12
5.1.1 Body Parts Module 12
5.1.2 Joint Module 12
5.2 Comparison of Optimizer Algorithms 12
5.3 Performance of loss functions 14
5.3.1 Sparse Categorical Cross Entropy 15
5.3.2 Mean Squared Error 15
5.3.3 Our custom loss function CILoss 15

6 Conclusion and future thoughts **17**

Acronyms

bl layer **L** with largest feature maps. xi, 5

laser A strange animal, not to be confused with. ix

List of Figures

5.1 Predicted Mask . . . 11
5.2 HRNet v7 . . . 12
5.3 Accuracy . . . 13
5.4 Correct BPR . . . 13
5.5 loss . . . 13
5.6 Learning Rate SGD . . . 14
5.7 Loss SGD . . . 14
5.8 Accuracy SGD . . . 14
5.9 CILoss . . . 15

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

1.1 Motivation and Goals

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.

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 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.

3.2 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

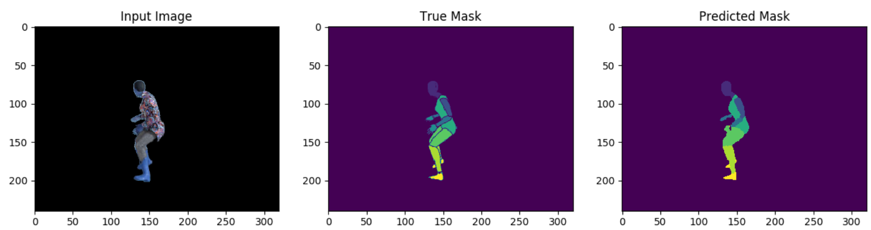

FIGURE 5.1: Predicted mask after 3845th epoch with custom loss function and Adam optimizer_kps

5.1 Ablation Study

5.1.1 Body Parts Module

Stride-down, -up convolution before bl

MobileNet extended with UNet

MobileNet extended with HRNet

Experiment with concat and add layers

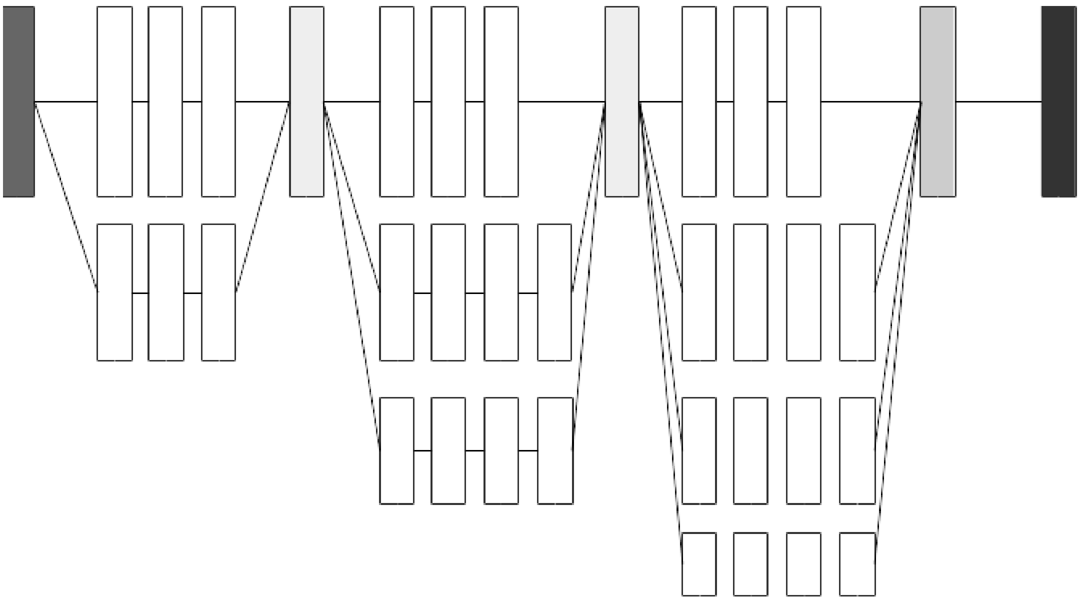

FIGURE 5.2: HRNet v7.

Best performing network HRNet 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

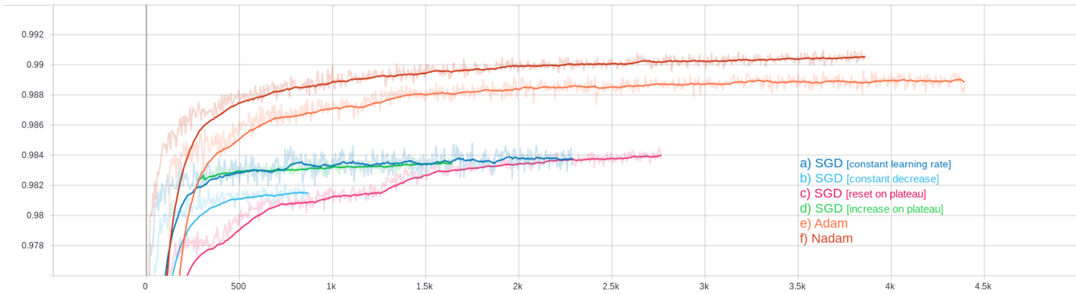

FIGURE 5.3: Accuracy

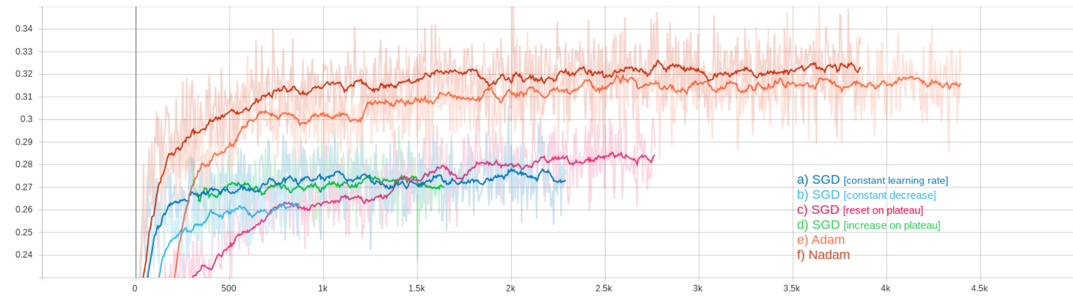

FIGURE 5.4: Correct body part pixel relation

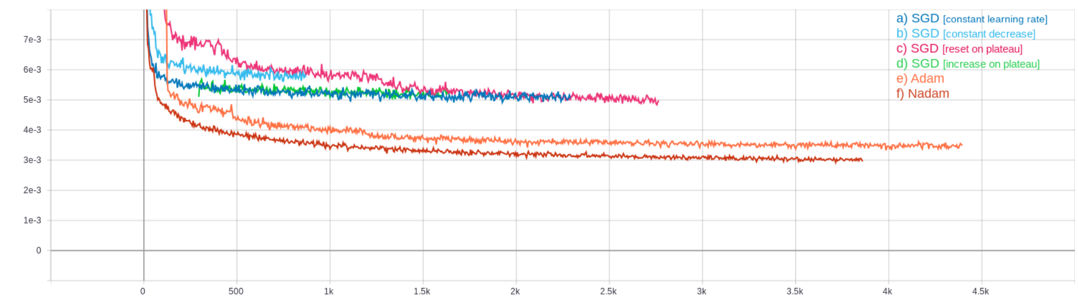

FIGURE 5.5: Loss

Comparison of Adam, Nadam and SGD

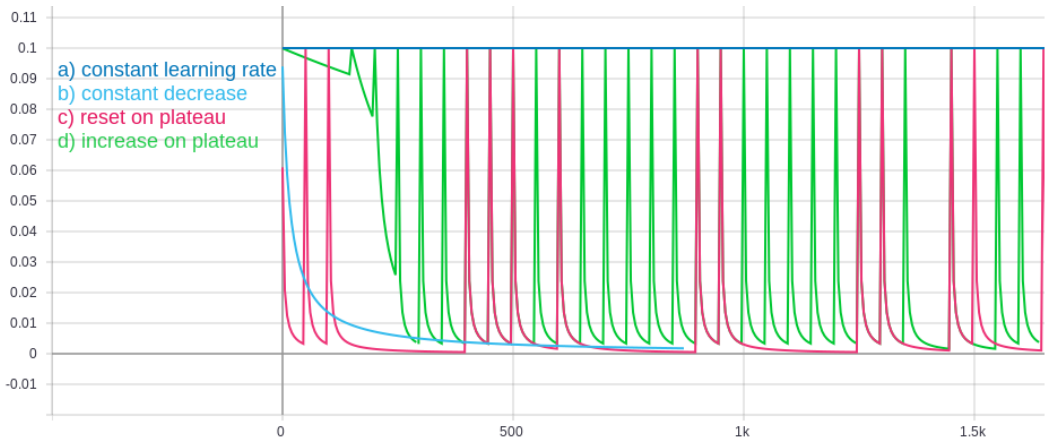

FIGURE 5.6: Learning Rate SGD.

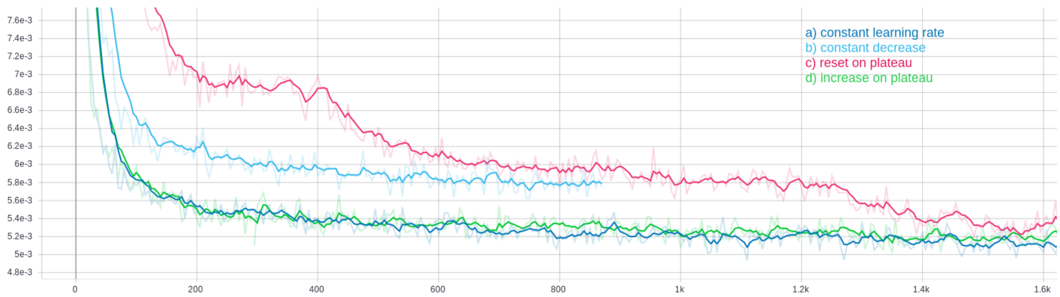

FIGURE 5.7: Loss SGD.

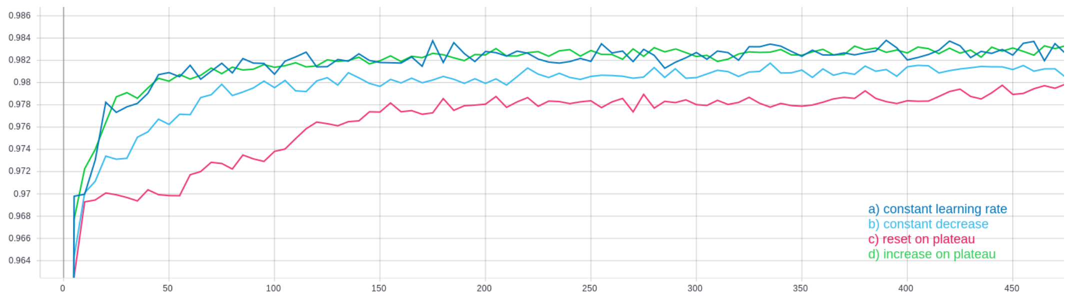

FIGURE 5.8: Accuracy SGD.

Experiments with SGD

5.3 Performance of loss functions

All performance measures are conducted on the Nadam optimizer_kps with the HR-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 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.

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

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

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

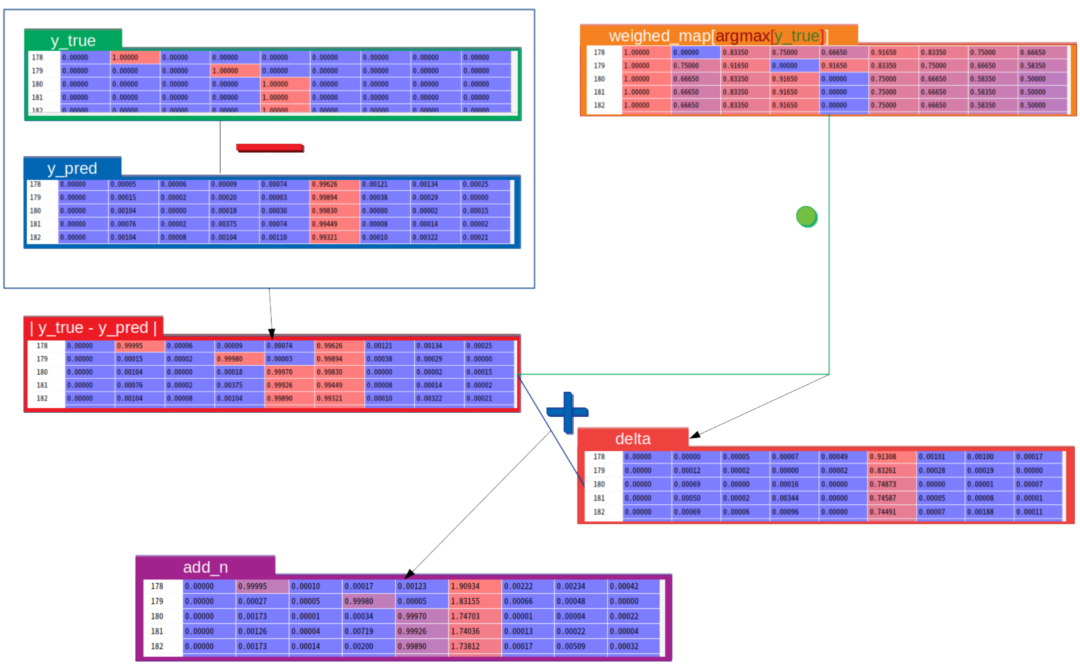

FIGURE 5.9: Visualization of custom loss calculation

Chapter 6

Conclusion and future thoughts