Thesis

Filip Kronström

January 19, 2021

Contents

Chapter 1

Introduction

skriv om risker med bias fr dataset osv...

1.1 Medical background

skriv om acl och varf;r detta arbete beh;vs related work, se ref i mendeley osv

1.1.1 POEs

...

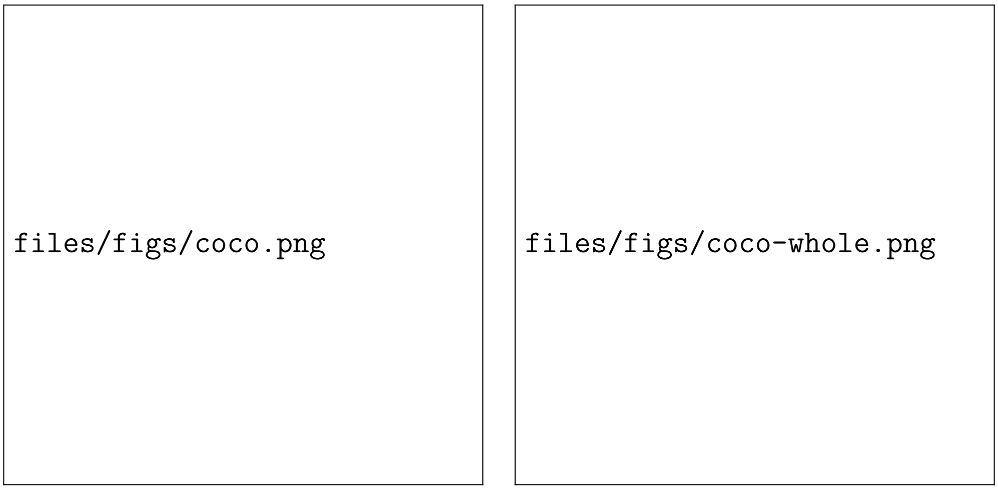
1.2 Machine learning

skriv om att ml blivit s[stort sen alexnet och gpu osv

The emergence of computing power discussed in Section 1.2 allowing deeper networks shown by AlexNet [9]

1.3 Pose estimation

Human pose estimation is a well explored problem which, like many other computer vision tasks has developed rapidly in the recent years. The reasons behind this progress can mainly be explained by two factors. Firstly the emergence of computing power discussed in Section 1.2, allowing more powerful deep learning models. Secondly several datasets with images labeled with human body joints has been made available [3]. These datasets



(a) Regular COCO. (b) COCO-wholebody.

Figure 1.3.1: Keypoints for the two COCO datasets.

not only provide data, but also introduces competition in the research community making it possible to compare the results of different approaches.

1.3.1 Datasets

Some of the widely used datasets today are Max Planck Institute for Informatics (**MPII**) [2], Microsoft Common Objects in Context (**COCO**) [10], AI Challenger Human Keypoint Detection (**AIC-HKD**) [15], and COCO-wholebody [8].

The COCO dataset consists of 328k images containing 91 different object types. The images come from Google, Bing, and Flickr image search and are mainly hand annotated through Amazon Mechanical Turk. The interesting part of the dataset for this work is the one with human poses. In total there are 250k instances of people labeled with joint locations [10]. The joints, 13 per person, in the dataset can be seen in Figure 1.3.1a. Along with the datasets containing body keypoints mentioned above there are also datasets with dense keypoints for specific bodyparts, e.g. OneHand10k [14]. COCO-wholebody is an attempt to combine these two types of datasets by extending COCO with dense keypoints at hands, feet, and faces. The resulting 133 joints can be seen in Figure 1.3.1b.

1.4 Time series classification

[6]

Chapter 2

Methods

2.1 Overview

In this chapter the system for assessing POEs will be presented. This system is naturally divided into two parts where firstly the videos are analyzed. The first subsystem extracts body part coordinates of the subjects. This information is then passed to the second subsystem where it is used to calculate a score according to [12]. The data is presented in Section 2.2 and the two subsystems are described in Sections 2.3 and 2.4 respectively.

2.2 Data

The data available is in the form of videos each containing one subject, recorded from the front, performing three-five repetitions of specific motions. The motions are *Single-leg squat, Forward lunge, Stair descending, Forward lunge, Single leg hop for distance, and Side hop*. Each motion has a number of POE scores associated with them. The motion-POE combinations are shown in Table 2.2.1. In Sections 2.2.1-**??** the motions and POEs evaluated in this project are described.

2.2.1 Single-leg squat, SLS

The subject performed a squat standing on one leg to a knee angle of approximately 60°. The exercise was repeated five times and the entire movement was used to assess the POEs [1]. An illustration is shown in Figure **??**.

Table 2.2.1: Motion-POE combinations available in the data.

POE \ Motion	Single leg squat	Stair descending	Forward lunge	Single leg hop for distance	Side hop
Trunk	x	x		x	x
Hip	x	x	x	x	x
Femoral valgus	x	x	x	x	x
Knee medial to foot position	x	x	x	x	x
Femur medial to shank	x	x	x	x	x
Foot	x				

2.2.2 Stair descending, SD

The subject stepped down from a 30 cm step board. The exercise was repeated five times and POEs were evaluated for the loaded leg during loading phase [1]. An illustration is shown in Figure ??.

2.2.3 Forward Lunge, FL

Helo I am now mispealing.

2.3 Body part localization

2.3.1 Preprocessing

... rotation, flip etc

2.3.2 Pose estimation

The pose estimation can be seen as a feature extraction and dimensionality reduction. The pose estimation is built around the open-source toolbox MMPose [11] from MMLab. Each frame is considered to be an independent image and is analyzed with the DARK-HRNet [13, 16] trained on the COCO-dataset [10] described in Section 1.3.

2.4 Classification

2.4.1 Preprocessing and dataset blabla..

Before assessing the POEs based on the body part positions a number of preprocessing steps are conducted. Firstly the data is resampled as the videos are recorded with a number of different frame rates (25, 30, and 60Hz). The resampling is performed using linear interpolation to a new sample frequency of 25Hz. This data is then low pass filtered through a fourth order Butterworth filter with a cutoff frequency of 2.5Hz. PLOT P[;VERF;RING F;R FILTER? ELLER P[FILTRERAD DATA?

While the POE assessment, see Section **??**, is performed on a per repetition basis the body part coordinates are extracted on a per video basis. Hence, each repetition should be extracted. This is done by finding peaks in the time series corresponding to certain body part positions. Which body part is used for this sequence splitting depends on which movement is analyzed. For SLS the y-coordinate of the right shoulder is used. The duration of each repetition varies significantly. With the duration of each repetition varying substantially between subjects padding in the time dimension is desirable. The reason for this is twofold, i) to simplify the handling of the data by storing it as a multidimensional array, and ii) to be able to train the eventual model in a more efficient manner using batches [5]. The padding is done by adding constant values of -1000 at the end of the sequences to some specified length. Details on how this is handled by the model are presented in Section **??**

BESKRIV ALGORITM F;R ATT DELA UPP I REPS

Finally the data is normalized to put the mean of the first five right hip-samples in the origin and the distance to the first five right shoulder-samples to one, according to (2.1).

$$
\begin{aligned}
(x,y)_i &= (x,y)_i - \overline{(x,y)}_{rh} \\
(x,y)_i &= \frac{(x,y)_i}{\|\overline{(x,y)}_{rs}\|_2} \qquad , \forall i
\end{aligned}
\tag{2.1}
$$

where: $\overline{(x,y)}_i$ = mean over first five samples for body part i
rh = right hip
rs = right shoulder
i $\in$ Available body parts

After these preprocessing steps a dataset $\in \mathbb{R}^{N\times T\times F}$ is created consisting of N multivariate time series with F channels of length T. The input channels can be x- and y-coordinates of body parts as well as angles between body parts.

2.4.2 Models

The proposed models are neural networks with convolutional layers as feature extractors. The network is inspired by InceptionTime by Fawaz et al. [7], described in Section **??**.

[4] lol

If you are using mendeley to manage references, you might have to export them manually in the end as the automatic ways removes the ”date accessed” field

Bibliography

[1] Jenny Älmqvist Nae. “Is seeing just believing? Measurement properties of visual assessment of Postural Orientation Errors (POEs) in people with anterior cruciate ligament injury”. English. PhD thesis. Department of Health Sciences, June 2020. ISBN: 978-91-7619-940-4.

[2] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. “2D Human Pose Estimation: New Benchmark and State of the Art Analysis”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 3686–3693. DOI: `10.1109/CVPR.2014.471`.

[3] Yucheng Chen, Yingli Tian, and Mingyi He. “Monocular Human Pose Estimation: A Survey of Deep Learning-based Methods”. In: *Computer Vision and Image Understanding* 192 (June 2020). DOI: `10.1016/j.cviu.2019.102897`. arXiv: `2006.01423`. URL: `http://arxiv.org/abs/2006.01423%20http://dx.doi.org/10.1016/j.cviu.2019.102897`.

[4] Kevin Fauvel, Tao Lin, Véronique Masson, Élisa Fromont, and Alexandre Termier. “XCM: An Explainable Convolutional Neural Network for Multivariate Time Series Classification”. In: (Sept. 2020). arXiv: `2009.04796`.

[5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[6] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre Alain Muller. “Deep learning for time series classification: a review”. In: *Data Mining and Knowledge Discovery* 33.4 (July 2019), pp. 917–963. ISSN: 1573756X. DOI: `10.1007/s10618-019-00619-1`. arXiv: `1809.04356`.

[7] Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre Alain Muller, and François Petitjean. “InceptionTime: Finding AlexNet for time

series classification". In: *Data Mining and Knowledge Discovery* 34.6 (Nov. 2020), pp. 1936–1962. ISSN: 1573756X. DOI: 10.1007/s10618-020-00710-y. arXiv: 1909.04939.

[8] Sheng Jin, Lumin Xu, Jin Xu, Can Wang, Wentao Liu, Chen Qian, Wanli Ouyang, and Ping Luo. "Whole-Body Human Pose Estimation in the Wild". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12354 LNCS (July 2020), pp. 196–214. arXiv: 2007.11858.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: ed. by Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger. 2012, pp. 1106–1114. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

[10] Tsung Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft COCO: Common objects in context". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8693 LNCS. PART 5. Springer Verlag, May 2014, pp. 740–755. DOI: 10.1007/978-3-319-10602-1_48. arXiv: 1405.0312.

[11] *MMPose - OpenMMLab Pose Estimation Toolbox and Benchmark*. URL: https://github.com/open-mmlab/mmpose.

[12] Jenny Nae, Mark W. Creaby, Gustav Nilsson, Kay M. Crossley, and Eva Ageberg. "Measurement Properties of a Test Battery to Assess Postural Orientation During Functional Tasks in Patients Undergoing ACL Injury Rehabilitation". In: *Journal of Orthopaedic & Sports Physical Therapy* 47.11 (Oct. 2017), pp. 1–42. ISSN: 0190-6011. DOI: 10.2519/jospt.2017.7270. URL: http://www.jospt.org/doi/10.2519/jospt.2017.7270.

[13] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. "Deep high-resolution representation learning for human pose estimation". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2019-June. IEEE Computer Society, June 2019, pp. 5686–5696. ISBN: 9781728132938. DOI: 10.1109/CVPR.2019.00584. arXiv: 1902.09212.

[14] Yangang Wang, Cong Peng, and Yebin Liu. "Mask-Pose Cascaded CNN for 2D Hand Pose Estimation from Single Color Image". In: *IEEE Transactions on Circuits and Systems for Video Technology* 29.11 (Nov. 2019), pp. 3258–3268. ISSN: 15582205. DOI: 10.1109/TCSVT.2018.2879980.

[15] Jiahong Wu, He Zheng, Bo Zhao, Yixin Li, Baoming Yan, Rui Liang, Wenjia Wang, Shipei Zhou, Guosen Lin, Yanwei Fu, Yizhou Wang, and Yonggang Wang. "AI Challenger : A Large-scale Dataset for Going Deeper in Image Understanding". In: *arXiv* (Nov. 2017). arXiv: 1711.06475. URL: http://arxiv.org/abs/1711.06475.

[16] Feng Zhang, Xiatian Zhu, Hanbin Dai, Mao Ye, and Ce Zhu. "Distribution-Aware Coordinate Representation for Human Pose Estimation". In: Institute of Electrical and Electronics Engineers (IEEE), Aug. 2020, pp. 7091–7100. DOI: 10.1109/cvpr42600.2020.00712. arXiv: 1910.06278.