FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Master Thesis

2D Tracking in Climbing

Using Temporal Smoothing

André Oskar Andersen (wpr684)
wpr684@alumni.ku.dk

2023

Supervisor
Kim Steenstrup Pedersen kimstp@di.ku.dk

Abstract

Preface

Acknowledgement

Contents

Notation

1 Introduction

- Imaged-based:
  - Geometry between joints in the target image:
    - Poselet conditioned pictorial structures
    - Exploring the spatial hierarchy of mixture models for human pose estimation
    - Articulated human detection with flexible mixtures of parts
  - Convolutional Pose Machine
  - Stacked Hourglass
  - OpenPose
  - HRNet

1.1 Related Work

2-dimensional pose estimation can be divided into either being image-based or video-based, where the methods in the latter case use the tempoeral information of the video to perform the pose estimation. Image-based methods [**MANGLER**]... . Video-based methods commonly use the correlating information among the frames of the video to perform the pose estimation. Early video-based methods used 3-dimensional convolutions to capture the correlating information between neighboring frames [5, 2]. Other methods use LSTM's [3] to capture the correlating information among the frames [4, 1]. Recently, transformers [6] have started to being used as a way of capturing the correlating information among the frames [7].

1.2 Problem Definition

1.3 Reading Guide

2 Deep Learning Theory

2.1 Feedforward Neural Networks

2.2 Convolutional Neural Networks

2.3 Recurrent Neural Networks

2.3.1 Long Short-Term Memory Unit

2.3.2 Gated Recurrent Unit

2.4 Transformer

3 Models

4 Dataset

5 Experiments

6 Discussion

7 Conclusion

8 References

[1] Bruno Artacho and Andreas Savakis. *UniPose: Unified Human Pose Estimation in Single Images and Videos*. 2020. DOI: 10.48550/ARXIV.2001.08095. URL: https://arxiv.org/abs/2001.08095.

[2] Rohit Girdhar, Georgia Gkioxari, Lorenzo Torresani, Manohar Paluri, and Du Tran. *Detect-and-Track: Efficient Pose Estimation in Videos*. 2017. DOI: 10.48550/ARXIV.1712.09184. URL: https://arxiv.org/abs/1712.09184.

[3] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[4] Yue Luo, Jimmy Ren, Zhouxia Wang, Wenxiu Sun, Jinshan Pan, Jianbo Liu, Jiahao Pang, and Liang Lin. *LSTM Pose Machines*. 2017. DOI: 10.48550/ARXIV.1712.06316. URL: https://arxiv.org/abs/1712.06316.

[5] Tomas Pfister, James Charles, and Andrew Zisserman. *Flowing ConvNets for Human Pose Estimation in Videos*. 2015. DOI: 10.48550/ARXIV.1506.02897. URL: https://arxiv.org/abs/1506.02897.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. *Attention Is All You Need*. 2017. DOI: 10.48550/ARXIV.1706.03762. URL: https://arxiv.org/abs/1706.03762.

[7] Ailing Zeng, Xuan Ju, Lei Yang, Ruiyuan Gao, Xizhou Zhu, Bo Dai, and Qiang Xu. *DeciWatch: A Simple Baseline for 10x Efficient 2D and 3D Pose Estimation*. 2022. DOI: 10.48550/ARXIV.2203.08713. URL: https://arxiv.org/abs/2203.08713.