Mendel University in Brno
Faculty of Business and Economics

Automated neural network learning for higher accuracy human skeleton detection under realistic conditions

Master's Thesis

Bc. Damián Sova

Supervisor:
doc. Ing. Oldřich Trenz, Ph.D.

Brno 2024

MENDELU
Faculty
of Business
and Economics

Department of Informatics
Academic year: 2023/2024

DIPLOMA THESIS TOPIC

Author of thesis:	**Damián Sova**
Study programme:	Open Informatics
Scope:	Scope for Open Informatics
Topic:	**Automated neural network learning for higher accuracy human skeleton detection under realistic conditions**
Length of thesis:	2,5–4 AA

Guides to writing a thesis:

1. The aim of this thesis is to analyze the problem of image detection using neural networks and to design a neural network model for human skeleton detection in real conditions.
2. Analyze the current state of the art in human skeleton detection, i.e., methods, approaches, available datasets.
3. Design a neural network model using available approaches (BodyPoseNet – NVIDIA, MediaPipe PoseNet – Google, ViTPose) for human skeleton detection. As part of the solution to this item, create a reference dataset (image dataset for skeleton detection).
4. Implement the proposed solution using freely available technologies. Use the Apple iPhone 14 Pro, Lidar, as a possible extension in the 3D level.
5. Evaluate your own solution and formulate options for further development.

Selected bibliography:

1. ARLOW, Jim; NEUSTADT, Ila. *UML 2 a unifikovaný proces vývoje aplikací: objektově orientovaná analýza a návrh prakticky.* 2nd ed. Brno: Computer Press, 2007. 567 p. ISBN 978-80-251-1503-9.
2. PATTON, Ron. *Software Testing.* Indiana: Sams Publishing, 2005. 408 p. ISBN 978-0-672-32798-8.
3. CHOLLET, François; PECINOVSKÝ, Rudolf. *Deep learning v jazyku Python: knihovny Keras, Tensorflow.* 1st ed. Praha: Grada Publishing, 2019. 328 p. Knihovna programátora. ISBN 978-80-247-3100-1.
4. CHOLLET, François. *Deep learning with Python.* Shelter Island: Manning, 2021. 478 p. ISBN 978-1-61729-686-4.
5. C. Patil and V. Gupta (2021, July 15). Human pose estimation using keypoint RCNN in pytorch. LearnOpenCV. https://learnopencv.com/human-pose-estimation-using-keypoint-rcnnin-pytorch/.
6. Rosebrock, A. (2021, April 17). R-CNN object detection with Keras, tensorflow, and Deep Learning. PyImageSearch. https://pyimagesearch.com/2020/07/13/r-cnn-object-detection-withkeras- tensorflow-and-deep-learning/.
7. Z. Tang, D. Wang and Z. Zhang, "Recurrent neural network training with dark knowledge transfer," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 5900-5904, doi: 10.1109/ICASSP.2016.7472809.

Diploma thesis topic submission date: May 2023

Deadline for submission of Diploma thesis: May 2024

L. S.

Electronic approval: 29. 6. 2023
doc. Ing. Oldřich Trenz, Ph.D.
Thesis supervisor

Electronic approval: 29. 6. 2023
Damián Sova
Author of thesis

Electronic approval: 29. 6. 2023
prof. Ing. Cyril Klimeš, CSc.
Head of Institute

Electronic approval: 29. 6. 2023
doc. Ing. František Dařena, Ph.D.
Study programme supervisor

I would like to thank the supervisor doc. Ing. Oldřich Trenz, Ph.D. for the time and provided advice. Next I would like to thank to the consultant RNDr. Michal Procházka, Ph.D. from visioncraft s.r.o. for regular consultantions which led the direction of this work. Last but not least, thank you friends and family for your support.

Declaration

I hereby declare that this thesis entitled *Automated neural network learning for higher accuracy human skeleton detection under realistic conditions* was written and completed by me. I also declare that all the sources and information used to complete the thesis are included in the list of references. I agree that the thesis could be made public in accordance with Article 47b of Act No. 111/1998 Coll., Higher Education Institutions and on Amendments and Supplements to Some Other Acts (the Higher Education Act), and in accordance with the current Directive on publishing of the final thesis. I declare that the printed version of the thesis and electronic version of the thesis published in the application Final Thesis in the University Information System is identical.

I am aware that my thesis is written in accordance to Act No. 121/2000 Coll. (the Copyright Act) and therefore Mendel University in Brno has the right to conclude licence agreements on the utilization of the thesis as a school work in accordance with Article 60 (1) of the Copyright Act.

Before concluding a licence agreement on utilization of the work by another person, I will request a written statement from the university that the licence agreement is not in contradiction to legitimate interests of the university, and I will also pay a prospective fee to cover the cost incurred in creating the work to the full amount of such costs.

Brno January 21, 2024 ....

signature

Abstract

Sova, Damián. *Automated neural network learning for higher accuracy human skeleton detection under realistic conditions.* Master's Thesis. Brno : Mendel University in Brno, 2024.

Key words

image processing, pose estimation, human skeleton detection

Abstrakt

Sova, Damián. *Automatizované učenie neurónových sietí na presnejšiu detekciu ľudskej kostry v reálnych podmienkach.* Diplomová práca. Brno : Mendelova univerzita v Brně, 2024.

Kľúčové slová

spracovanie obrazu, odhadovanie pózy, detekcia ľudskej kostry

Contents

1 Introduction

Motivation and basic objectives of the work
Current state and problem to be addressed

2 Theoretical foundations

This chapter provides an overview of the theoretical foundations of the proposed automated neural networks (NN) learning approach for human skeleton detection. It introduces the key concepts of NNs, convolutional neural network (CNN), region-based convolutional neural network (RCNN), and transformation models of NNs. Additionally, it explores existing NNs for human pose detection, including Posenet, Movenet, and MMPose.

2.1 Neural Network

NNs are computational models inspired by the structure and function of the human brain. They consist of interconnected layers of artificial neurons that process and transform information. NNs have demonstrated remarkable capabilities in various tasks, including image recognition, natural language processing, and machine translation.

2.2 Convolutional Neural Network

CNNs are a type of NN architecture that excels at processing and analyzing visual data, such as images and videos. They are particularly well-suited for skeleton detection due to their ability to extract local features from the input data. CNNs typically consist of a series of convolutional layers, each of which applies a filter or kernel to the input data to extract features. The filters are learned during the training process, allowing the CNN to learn the patterns and relationships that are important for skeleton detection.

CNNs have several advantages for skeleton detection:

Translation Invariance: CNNs are invariant to small translations in the input data. This is important for skeleton detection, as the human body can be in different positions in an image or video.

Feature Learning: CNNs can learn complex features from the input data, which is essential for accurate skeleton detection.

Parameter Sharing: CNNs share weights across different positions in the input data. This reduces the number of parameters in the network, making it more efficient and easier to train.

CNNs have become the dominant architecture for skeleton detection, and they have significantly improved the accuracy of this task.

2.3 Region-based Convolutional Neural Network

RCNNs are a type of NN architecture that excels at object detection and localization. They typically follow a two-stage process:

- **Region Proposal:** An initial set of regions of interest (ROI) is proposed in the input image.
- **Feature Extraction and Classification:** For each ROI, a feature vector is extracted using a CNN and classified as containing the object or not.

2.4 Transformation Models of NNs

Transformation models aim to improve the performance and efficiency of NNs by transforming the input or output data. These models can be used to reduce the dimensionality of the data, improve the interpretability of the model, or adapt the model to specific tasks.

2.5 Existing NNs for Human Pose Estimation

Several NN architectures have been developed for skeleton detection. Here are three notable examples:

(1) **Posenet** (Mediapipe): A lightweight and efficient NN for human pose estimation. It uses a single-stage architecture and can run on mobile devices.

(2) **Movenet** (TensorFlow): A multimodal NN that combines pose estimation, hand tracking, and object tracking. It offers a variety of models with different tradeoffs between accuracy and speed.
(3) **MMPose** (Open-MMLab): A modular and extensible library for pose estimation. It provides a wide range of models and training tools.

2.6 Chapter Summary

This chapter introduced the key concepts of NNs, RCNNs, transformation models, and existing NNs for human pose estimation. These concepts provide the theoretical foundation for the proposed automated NN learning approach.

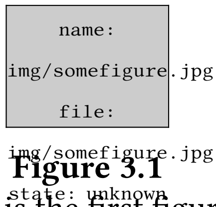

Figure 3.1
This is the first figure.

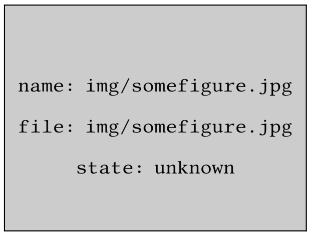

Figure 3.2
This is the second figure.

3 Practical part

Dataset description
Description of the created data format for each stage of the training process
Description of the impementation experiments and results
Implementation problems and technical limitations
Some text referring to **Figures 3.1 and 3.2**.

4 Conclusion

Evaluation of the achieved results
Suggestions for further improvements
Summary of the results of the work

References

This is the room for references, learn BIB.STY, the style for references.

List of Tables

List of Figures

List of Abbreviations

CNN	Convolutional neural network
NN	Neural network
RCNN	Region-based convolutional neural network
ROI	Regions of interest

List of Source Codes

APPENDICES