Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Comparing Synthetic and Real Data for Anthropometric Measurements Estimation

Bachelor Thesis

2024

Michal Baránek

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Comparing Synthetic and Real Data for Anthropometric Measurements Estimation

Bachelor Thesis

Study Programme:	Computer Science
Field of Study:	Computer Science
Department:	Department of Computer Science
Supervisor:	Mgr. Dana Škorvánková

Bratislava, 2024
Michal Baránek

18794588

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Michal Baránek
Študijný program:	aplikovaná informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor:	informatika
Typ záverečnej práce:	bakalárska
Jazyk záverečnej práce:	anglický
Sekundárny jazyk:	slovenský

Názov: Comparing Synthetic and Real Data for Anthropometric Measurements Estimation
Porovnanie syntetických a reálnych dát pre účely odhadu antropometrických mier

Anotácia: Odhad mier ľudského tela je úloha, ktorá priťahuje v posledných rokoch pozornosť viacerých vedeckých oblastí. Automatický a presný prístup na riešenie tohto problému je kľúčový v rôznych oblastiach počítačového videnia. Výroba odevov a šitie odevov na mieru patria medzi aplikácie, kde by presný odhad mier tela z vizuálnych dát človeka bol prínosným, nahradením tradičného manuálneho merania tela.

Cieľ: Cieľom tejto bakalárskej práce je naštudovať problematiku odhadu rozmerov ľudského tela, otestovať vybrané state-of-the-art metódy založené na hlbokom učení a evaluovať ich pomocou syntetických aj reálnych dát. Úlohou je týmto spôsobom analyzovať rozdiely medzi danými doménami, a vyhodnotiť výhody augmentácie reálnych dát pomocou syntetických obrazov pre účely trénovania modelu.

Kľúčové slová: miery tela, neurónové siete, hlboké učenie

Vedúci:	Mgr. Dana Škorvánková
Katedra:	FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry:	doc. RNDr. Tatiana Jajcayová, PhD.
Dátum zadania:	25.09.2023
Dátum schválenia:	04.10.2023

doc. RNDr. Damas Gruska, PhD.
garant študijného programu

.... študent

.... vedúci práce

18794588

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Michal Baránek
Study programme:	Applied Computer Science (Single degree study, bachelor I. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Bachelor´s thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Comparing Synthetic and Real Data for Anthropometric Measurements Estimation

Annotation: Human body measurements estimation is a task that attracts the attention of several scientific fields in recent years. An automatic and accurate approach to this problem is crucial in various areas of the computer vision-oriented industry. Garment manufacturing and tailoring are some of the applications, where an accurate body measurements estimation from visual human data would be beneficial, replacing the traditional manual tape measuring.

Aim: The goal of the bachelor thesis is to study the task of body measurements estimation. The aim is to test selected state-of-the-art deep learning methods and evaluate them using both synthetic and real human body data. In this way, we aim to analyze the domain gap and explore the benefits of augmenting real data with synthetic images for training purposes.

Keywords: body measurements, neural networks, deep learning

Supervisor:	Mgr. Dana Škorvánková
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	doc. RNDr. Tatiana Jajcayová, PhD.
Assigned:	25.09.2023

Approved: 04.10.2023 doc. RNDr. Damas Gruska, PhD.
Guarantor of Study Programme

....

Student Supervisor

Acknowledgments: Tu môžete poďakovať školiteľovi, prípadne ďalším osobám, ktoré vám s prácou nejako pomohli, poradili, poskytli dáta a podobne.

Abstrakt

Slovenský abstrakt v rozsahu 100-500 slov, jeden odstavec. Abstrakt stručne sumarizuje výsledky práce. Mal by byť pochopiteľný pre bežného informatika. Nemal by teda využívať skratky, termíny alebo označenie zavedené v práci, okrem tých, ktoré sú všeobecne známe.

Kľúčové slová: jedno, druhé, tretie (prípadne štvrté, piate)

Abstract

Abstract in the English language (translation of the abstract in the Slovak language).

Keywords:

Contents

List of Figures

List of Tables

Introduction

Nowadays our physical bodies are not enough for us anymore. Thanks to modern technologies every day we come closer to living a new type of life. Virtual one. The beginnings were quite humble with us sharing our thoughts via text and shortly after empowering our life stories with images. Thanks to these innovations we were able to create a new type of business - online shops. Their popularity rose over the years and are now rivals to traditional ways of shopping. These new online stores however have a significant disadvantage in comparison. You cannot try the clothes on and see whether it will fit. Luckily we have already created clothing sizing system to help us choose the correct size. However to use this system one would need exact body measurements to look up correct size in the table.

Online shops however were not the only blooming business in recent years. A new prospect has appeared only just recently and it caught attention of big companies straight away. The virtual world where we could meet our friends living anywhere in the world without having to leave our home. In this world everyone would have their own avatar that would represent them. Some virtual worlds would let you become whoever or even whatever you could possibly think of while some others decided to stick with realism. Realism sounds great, but this approach requires realistic data to be able to look convincing. To create a clone of human body in the virtual world we amongst many variables would need exact body measurements to provide a believable result.

These are just two examples that require users to obtain their body measurements. While the act of measuring does not seem very problematic the measurements are prone to human error. There are no rules when it comes to measuring body parts. Usually, subjects are only guided via text or image showing them how to measure. This will never satisfy the accuracy that is required.

Thanks to progress in neural networks a new way of obtaining body measurements has emerged. With requiring only picture of body from front we can train a neural network to predict measurements of human body. This has already been proven plausible, but the accuracy lacked in some measurements. The goal of this work is to compare different .... (To make data better). The methods will vary to see how impactful they are regarding the final result.

For training we will use a synthetic dataset which will allow us to have much larger number of samples for training. For metrics we will use performance on dataset with real samples provided by BodyM dataset.

The first chapter will provide overview of the problematic, will look into traditional measuring methods, delve into obstacles this approach faces and then explain some of the mechanical works of this thesis. Second chapter will look into already existing work that is relevant to topic. In the third chapter we will define datasets. Contents of fourth chapter will focus on proposed solution and implementation. Results of our research will be located in fifth chapter while the sixth chapter will provide conclusion.

Chapter 1

Overview

1.1 Traditional measuring methods

Human body can be measured using many different approaches. All of these have their advantages and disadvantages. In this thesis we only need to introduce two of them.

Hand measurement

This is the traditional method of using tape measure for obtaining measurements. This approach usually needs one extra person that performs the measurement on the subject. This way the measurements are more precise than the subject using the tape measure without help. The measurements however have to be taken at specific locations to provide correct information in further processing. The locations vary depending on the use and thus there is no universal guide for their measurement. Whereas the professionals are familiar with them, the subjects are usually not as informed. This creates space for human error we are trying to avoid.

Digital measuring on 3D model

This method is mentioned because BodyM dataset uses this method to provide ground truth values. Is it based on photogrammetry scanner which creates precise meshes of scanned object. After the human body get acquired from the scanner they are registered to SMPL [Loper et al., 2015] mesh topology. We can then use the resulting mesh for calculating the measurements.

1.2 Obstacles

One of the issues when working with human body measurements is the lack of real world data. The process of measuring is time-consuming and requires privacy measures to

take place to protect subjects' personal information. This can be avoided by using synthetic datasets.
Moreover the time complexity to train a neural network can be reduced by using powerful device which is not available.

1.3 Problem specification

The goal of this thesis is to expand on body measurements estimation with the use of neural networks. Mainly to explore different data augmentation methods to provide better results when training is done on synthetic dataset.

1.4 SMPL

SMPL is short for Skinned Multi-Person Linear model. This model is based on skinning and blend shapes. The model has been trained on thousands of 3D body scans to create a realistic human model. Full explanation of functions and workflow of this model is out of scope of this work.

1.5 Neural network

Neural network is code built on premises of how human brains work. It consists of connected nodes called neurons. Each neuron takes input variables, processes them and then sends the result to other neurons. Every connection has associated weight which determines the influence said value will have. Neurons are then organised into layers. They are usually divided into input, output and hidden layers. The function of the hidden layers is to perform the operations needed to calculate output from the input data. The process of training adjusts the weights of the connections. This automatic process of adjusting is usually based on comparing the output and correct value we provide for the network and minimising the difference. This process helps the network to find complex relationships or patterns that may not be as understandable for humans.

Hyperparameter

Hyperparameter is a parameter that is not learned but chosen by developer. These parameters do not change over time. These can be - choice of optimizer, learning rate, number of layers, filter size and more.

Our model is mainly built on the following layers:

Convolution layer

Most popularly used with convolutional neural networks [O'Shea and Nash, 2015] this layer plays important role in network's functionality. It is based on working with matrices called kernels. The values in kernel are learnable, which means they are adjusted over the training process to enhance performance. In this thesis we will be using these hyperparameters:

- **Depth** determines dimension of the output volume (activation maps). Influences pattern recognition as well as number of neurons.
- **Size** determines dimensions of kernel.
- **Activation function**

The algorithm consists of sliding kernel along the input. At every position it calculates the sum of element-wise multiplication of corresponding pixels in input and kernel. The result is then inserted into the output. This process is then repeated over the whole input multiple times (depending upon number of kernels) Result of this operation captures local patterns while preserving positional relationships.

Max Pooling layer

Max pooling is an operation of non-linear down-sampling. This means that the output image of this layer is usually smaller than the input. This helps to reduce parameters for next convolutional layer, providing faster training. This layer is defined by two hyperparameters:

- **Filter size** determines the dimensions. In case of the filter reaching out of the array, only valid values are taken into consideration.
- **Stride** determines how many columns will the filter move.

The higher these hyperparameters' values are the smaller will the output be.
This layer iterates over input field and looks at subfield with size of filter size. In this subfield it finds the largest number and writes only the largest number into the output field. After this, the filter moves by stride columns left until all columns were checked. In that case the filter moves back to first column of the input field and then moves down by the stride (refer to 1.1 as an example).
This process is repeated until whole input field is iterated.

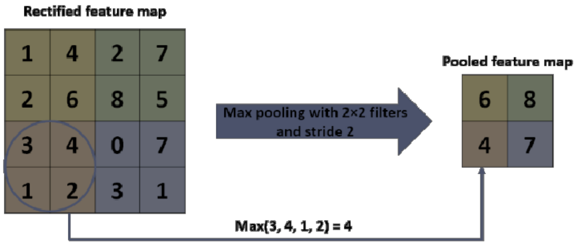

Figure 1.1: Max Pooling Example [Gholamalinezhad and Khosravi, 2020]

Overfitting

1.6 Keras [Chollet et al., 2015]

Keras is an open-source neural network library written in Python. Thanks to its user-friendly interface and modular design is Keras one of the leading frameworks in neural network development. Its simple yet flexible architecture allows for easy prototyping and experimentation, making it an ideal choice for both beginners and experienced practitioners in the field of deep learning.

1.7 OpenCV [Bradski, 2000]

Open Source Computer Vision Library (OpenCV for short) is a comprehensive open-source library originally developed by Intel. It is mainly used for various tasks in fields such as computer vision or machine learning. In the time of writing this thesis, OpenCV provides over 2500 optimized algorithms. These are able to effectively perform many tasks such as face detection, object tracking, image preprocessing and many more. Providing interfaces in multiple programming languages such as Python, C++, Java and MATLAB it is very popular with community as well as recognisable and famous companies.

1.8 Photogrammetry

Chapter 2

Related Work

There have already been multiple attempts to estimate human body measurements from a single image. The solution proposed in [BenAbdelkader and Yacoob, 2008] provides us with a method which estimates subject's height using single uncalibrated image. Another approach [Bogo et al., 2016] uses joints position estimation to create a 3D mesh used for further evaluation.

2.1 Neural Anthropometer

An important article is the [Tejeda and Mayer, 2021] which proposes a method to tackle this task. Its Neural Anthropometer provides a valuable approach which will we use as backbone our convolutional neural network architecture. We do not need everything used in this article as we already have annotated synthetic dataset provided by [Skorvankova et al., 2021]. To keep the network as small as possible due to resource consumption and training difficulty increase with size. The proposed architecture starts with a binary image silhouette input. This is then processed by a convolutional layer. Number of channels was based on number of values on output. The output tensor is then passed through ReLU [Nair and Hinton, 2010] along with batch optimization. Subsequently, a max pooling layer is used followed by a convolutional layer. The output is then once again processed by max pooling layer with configuration same as before. The result is then flattened to a tensor which is passed to a fully connected layer and a ReLU. As the last layer a regressor is used to provide the measurements estimation.

Since this thesis isn't about tweaking architecture for better results, but rather, it's all about data augmentation, and this architecture gives us the effectiveness and precision sufficient for demonstration purposes, we will build upon this architecture.

Chapter 3

Dataset overview

In this section, we'll explore the datasets utilized within this thesis. Our focus will be on 2D front-facing and profile human binary silhouettes. This form was chosen upon the data provided by the BodyM dataset. The subjects are positioned in an a-pose, ensuring greater consistency in the samples.

3.1 SURREACT

Description

SURREACT [Varol et al., 2021] is a synthetic dataset built on SMPL model. The main goal of the work was to explore benefits of using synthetic data for human action recognition. The study aimed to answer whether the synthetic data could potentially improve accuracy of already existing methods. This theory was confirmed and even shown improvements over other state-of-the-art action recognition methods. This is however not as important for this thesis as we are not going to use the features that were added.

The dataset introduced by [Skorvankova et al., 2021] is an extension of the SURREACT dataset, incorporating the data generation techniques and a custom annotation method. This thesis utilizes a modified version of this dataset. The original dataset comprises 50,000 human scans, meshes, annotations, and other data of subjects in the T-Pose. In contrast, our customized version offers 79,999 frontal and 79,999 lateral images with annotations, featuring subjects in the A-Pose. They are saved in RGBA format with dimensions of 320x240 without background thus eliminating the need of segmentation. Measurements are saved in .npy file format requiring us to use NumPy [Harris et al., 2020] to read these values.

Measurements

The measurements provided by this dataset are rather clearly described.

SMPL

3.2 BodyM

Description

This public body measurement dataset [Ruiz et al., 2022] contains measurement and image data from real human subjects. The subjects were photographed in a well-lit indoor setup, resulting in the data being less prone to segmentation inaccuracies. Subjects also wore tight-fitting clothing to better reflect the measurements. After the pictures were taken, the subjects were scanned using Treedy photogrammetric scanner and fitted to the SMPL mesh. Measurements were then taken on said meshes.

Measurements

Issues

Chapter 4

Development

4.1 Suggested approach

4.2 Setting up network

4.3 Figuring out measurements

4.4 Optimalizations

4.4.1 Including BodyM in training

4.4.2 Adding height

4.4.3 Adding lateral image

4.4.4 Combinations

Chapter 5

Results

Chapter 6

Conclusion

Bibliography

[BenAbdelkader and Yacoob, 2008] BenAbdelkader, C. and Yacoob, Y. (2008). Statistical body height estimation from a single image. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–7.

[Bogo et al., 2016] Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., and Black, M. J. (2016). Keep it smpl: Automatic estimation of 3d human pose and shape from a single image.

[Bradski, 2000] Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

[Chollet et al., 2015] Chollet, F. et al. (2015). Keras. `https://github.com/keras-team/keras`.

[Gholamalinezhad and Khosravi, 2020] Gholamalinezhad, H. and Khosravi, H. (2020). Pooling methods in deep neural networks, a review.

[Harris et al., 2020] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

[Loper et al., 2015] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., and Black, M. J. (2015). SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16.

[Nair and Hinton, 2010] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

[O'Shea and Nash, 2015] O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks.

[Ruiz et al., 2022] Ruiz, N., Bellver, M., Bolkart, T., Arora, A., Lin, M. C., Romero, J., and Bala, R. (2022). Human body measurement estimation with adversarial augmentation.

[Skorvankova et al., 2021] Skorvankova, D., Riečický, A., and Madaras, M. (2021). Automatic estimation of anthropometric human body measurements.

[Tejeda and Mayer, 2021] Tejeda, Y. G. and Mayer, H. A. (2021). A neural anthropometer learning from body dimensions computed on human 3d meshes. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8.

[Varol et al., 2021] Varol, G., Laptev, I., Schmid, C., and Zisserman, A. (2021). Synthetic humans for action recognition from unseen viewpoints. In *IJCV*.