Predicting Human Body Dimensions from Single Images: a first step in automatic malnutrition detection

Hezha MohammedKhan[1,2], Marleen Balvert[1], Cicek Guven[2], and Eric Postma[2]
{H.H.Mohammedkhan@tilburguniversity.edu, C.Guven@tilburguniversity.edu, M.Balvert@tilburguniversity.edu}
[1] Zero Hunger Lab, Dept. of Econometrics & Operations Research, TISEM, Tilburg University, The Netherlands
[2]Cognitive Science & AI, TSHD, Tilburg University, The Netherlands

Abstract. Malnutrition in children accounts for 45% of child deaths globally. Automatic malnutrition detection from digital photos serves as a decision support tool for early detection of malnutrition in rural areas. We study the feasibility of estimating body-shape characteristics from images of human body shapes as a first step in automatic malnutrition detection. We generate multi-view images of male and female bodies from rendered digital 3D scans of human bodies. Using convolutional neural networks (CNNs), we estimated waist circumference and body height with a mean absolute error of 59 mm and 9 mm, respectively. The estimation error of waist circumference depends on viewpoint. We conclude that automatic malnutrition detection from single images seems feasible, provided one or more suitable viewpoints are used.

Keywords: convolutional neural networks, hunger, malnutrition, human body shape.

1 Introduction

The United Nations Sustainable Development Agenda for the year of 2030 states hunger as one of the world's most critical issues.[1] Of particular concern is malnutrition in children, prevalent in the less developed parts of the world [1]. Currently, more than 230 million children are malnourished [2, 3]. Early diagnosis of malnutrition in children allows actions to limit its negative impact on health and development. Traditional diagnosis mainly relies on specific body measurements such as length and weight of a child in relation to age, as well as the middle-upper arm and head circumference. Unfortunately many children do not have access to health services to be checked and diagnosed for malnutrition. In the absence of medical centers - which would ideally be the primary point for monitoring healthy growth of a child - health organizations, parents and medical workers would benefit from an easy-to-use app that detects signs of malnutrition from photos or video sequences to

[1] https://sustainabledevelopment.un.org/

CAIP 2021, November 20-24, Bologna, Italy

DOI 10.4108/eai.20-11-2021.2314166

alert health workers and parents. So the primary role of the app is to raise a flag where there is risk for follow up actions, but not to replace human intervention.

Our research is inspired by Welthungerhilfe's Child Growth Monitor[2], a mobile app under development that detects malnutrition in children under the age of five. In the context of the collaboration of our Zero Hunger Lab with Welthungerhilfe, we focus on exploring the feasibility of various approaches to malnutrition detection from single photos. In this paper, we report on our initial experiments to determine the feasibility of inferring body-shape characteristics from digital scans of humans with convolutional neural networks.

Essentially, our task is a body-shape estimation task. In the context of human body reconstruction, a recent study [4] distinguished model-based and model-free methods for body shape estimation. Model-based methods rely on a parameterized human body model such as SMPL [5] and try to estimate the parameter values from images. Model-free methods do not rely on a pre-defined model and estimate the body shape directly. Our approach is model free and examines to what extent two different convolutional neural networks (CNNs), namely Resnet [6] and Inception [7], are able to estimate body shape characteristics from single images. One of the major obstacles for studying human body shapes is the scarcity of publicly available data. Obviously, data on children's body shapes is difficult to obtain due to privacy and ethical concerns. Anticipating our own collection of data on the body shapes of children, we studied adult body shapes by using publicly available data on adult body shapes. We assume that our results obtained on adult body shapes generalize well to the body shapes of young children.

1.1 Related work

Several other studies developed neural networks for body shape estimation from full body images [8] or videos [9]. Most notably, Zheng et al. [10] proposed a method called SHARP for shape-aware reconstruction of people in loose clothing from a single image. Their end-to-end trainable network accurately recovers the geometry and appearance of human bodies in loose clothing. They used a body-shape prior based on SMPL [5] to link visual cues in 2D to depth estimates. The results of SHARP and similar approaches strengthen our expectation that inferring body-shape characteristics from a single image is possible. Since our task is more restricted than full-body reconstruction, we will not use a body-shape prior and examine to what extent our CNNs can exploit 2D cues to infer 3D characteristics.

Noteworthy are two studies [11] and [12] that both employ CNNs to predict body shape from images. Dantcheva, Bremond and Bilinski [11] trained a Resnet on facial images to predict the Body Mass Index, which is normally computed from the mass and height of a person. Dibra et al. [12] used a customized CNN architecture on silhouettes of the CAESAR data set, using restricted viewpoints.

[2] https://childgrowthmonitor.org/

2 Methods

In this section we describe the dataset used in our experiments, the processing pipeline, and the experimental setup.

2.1 Dataset

The acquisition of images and body-shape characteristics of children at risk of malnutrition in less developed regions poses ethical and practical obstacles. Given the generic nature of human bodies, we decided to start with a subset of pre-processed scans of the CAESAR dataset without additional posture normalization, made publicly available by Yang et al. [13, 14][3]. The data consists of meshes of 1531 female and 1518 male body shapes.

2.2 Processing pipeline

Figure 1 provides a schematic overview of the processing pipeline to run our experiment. Given the 3D meshes, we perform two operations to obtain inputs and associated labels to train our CNNs. The two operations are represented by the two oriented arrows in the figure. First, for each gender we perform principal component analysis on all meshes (cf. [13, 5]) to obtain body-shape labels, the first 10 principal component values, that are needed for supervised training of the CNNs. For each gender we selected the two principal components that capture length (PC1) and waist circumference (PC3 and PC4 for females and males, respectively). Second, for each body shape we render 2D-view images from 100 different viewpoints using Blender 2.82[4]. The viewpoints are defined as 100 equally-spaced viewpoints on the circle in the horizontal plane with a radius of 1 meter surrounding the middle of the body. The virtual camera is positioned at 1 meter height. Viewpoint 0 corresponds to the frontal image of the body. Combining the rendered images as inputs and the principal component values as labels, we train a CNN. After training, we validate the CNN on previously unseen body shapes by determining the mean absolute error (MAE) translated into millimeters. The translation was based on the ratio of the average value of PC1 and the average height as is publicly available for the CAESAR dataset.

2.3 Experimental setup

We used Keras/Tensorflow to implement our experiment. As CNNs we selected either the ResNet [6] or the Inception V3 [7] architecture. In both cases the CNNs were pretrained on Imagenet. Training (transfer learning) was performed on a single NVIDIA GeForce 960MX with 8 GB memory, core i7 8th generation processor and 16GB RAM. For both CNNs we replaced the last classification layer by a regression layer. To this end we used two linear activation functions to predict both PC values. The learning rate was set to 1e-3 initially, with a "reduce learning rate on plateau" scheduler with a reduction factor of 0.1. The batch size was set to 32.

[3] http://humanshape.mpi-inf.mpg.de/

[4] https://www.blender.org/

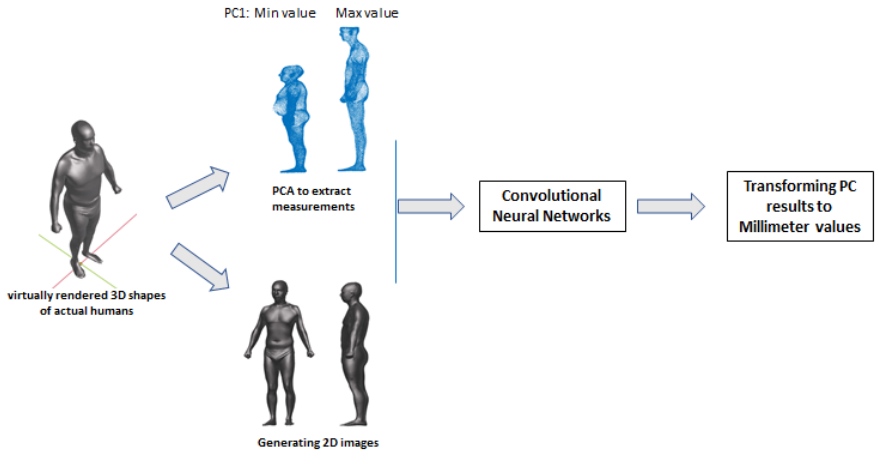

Fig. 1. An overview of our pipeline to extract labels (top) and inputs (bottom) from 3D meshes of actual human bodies.

We performed separate experiments for the male and female body shapes and for both CNNs. Our training, validation, and test sets consisted of 70%, 10%, and 20% of the body shapes, respectively.

3 Results

Before turning to the main results, Figure 2 illustrates the results of the principal component analysis of the female body-shape meshes. The figure shows the extreme values of five principal components. The value of the first component (PC1) represents height and the value of the third component (PC3) represents waist circumference.

We now turn to the main results. We start by presenting the results for the prediction of height, followed by those of weight circumference.

3.1 Results height prediction

Table 1 presents the results of height prediction. As can be seen from the results, the Inception model is outperforming the ResNet model, with an MAE of 9 mm for Inception versus an MAE of 10 mm for ResNet. There is a noticeable difference between the male and female results: both models perform better for males than for females. Inception has an MAE of 7 mm for males versus 11 mm for females. Our results compare favourably with those of [11] and [12] who reported MAEs of 78 mm and $10-12$ mm, respectively.

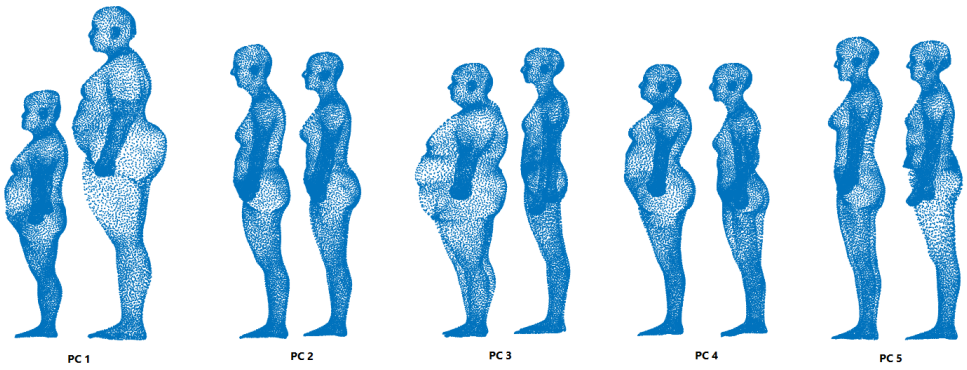

Fig. 2. Visualization of the extreme values of the first five principal component for the female data set.

Table 1: Performances of the two CNNs on height prediction.

CNN Model	MAE female in mm	MAE male in mm	Average MAE in mm
ResNet	12	8	10
Inception	11	7	9

3.2 Results waist circumference prediction

Table 2 presents the results obtained from training the CNN models on waist circumference. Similar to the prediction of height, the Inception model is slightly ahead of the ResNet model. The average MAE for Inception is 59 mm versus 60 mm for ResNet. The prediction for the male data set significantly outperformed the prediction for the female dataset for both models: Inception shows an MAE of 44 mm for males and an MAE of 73 mm for females.

Table 2: Performances of the two CNNs on waist-circumference prediction.

CNN Model	MAE female in mm	MAE male in mm	Average MAE in mm
ResNet	76	44	60
Inception	73	44	59

Figure 3 provides an illustration of the viewpoint dependency of the estimates of a trained CNN (Inception V3) for the waist circumference of the females. The horizontal axis represents the viewpoint ranging from 0 frontal image of the body, via 50 image of the back of the body, back to near frontal 100. The vertical axis represents, for each viewpoint, the MAE averaged over the body shapes. The box whiskers plots provide for each viewpoint an impression of the variation. The main observation of this graph is that the estimates vary with viewpoint and that there is quite some variation over body shapes. This variation is a bit smaller for the front (viewpoint 0 and 100) and to a lesser extent the side (around viewpoint 50).

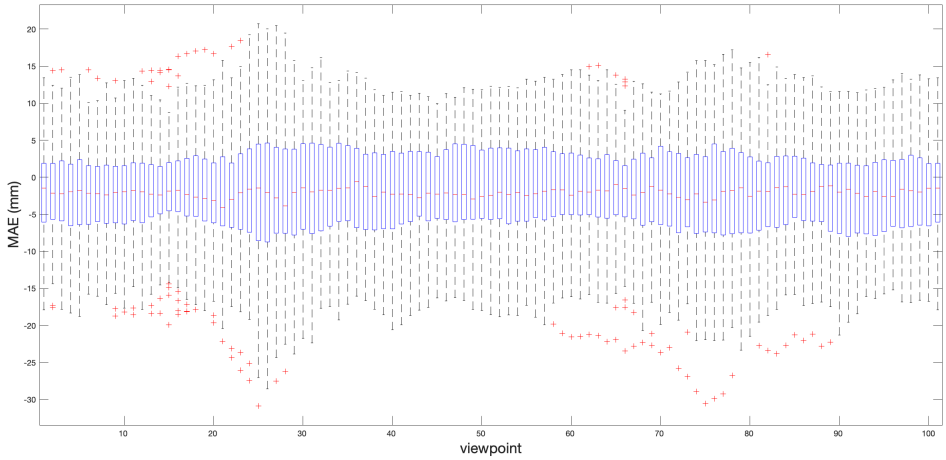

Fig. 3. Box whisker plots of view-dependent MAEs. Each boxplot shows the distribution over all female body shapes for a particular view.

3.3 Discussion of results

How good are our predictions? To put our height prediction results in context, we consider the results reported in the literature. The MAEs on height prediction reported in the literature range from 73 to 78 mm [11, 15, 16]. These MAEs are much larger than ours (more than 60 millimeters) but this may reflect the relatively standardized nature of our dataset. The best performing method in the literature uses full body images from surveillance cameras [16] and achieves an MAE of about 14 mm which is closer to ours.

Our results for waist circumference prediction are less accurate than those for height prediction. Presumably this reflects the fact that inferring waist circumference is much harder than inferring height of a body against a uniform background. The former requires the use of subtle shape cues, such as shape from shading, whereas the latter can be done without such cues. Imposing constraints on the viewpoint, i.e., the angle from which a body is photographed, will certainly improve the prediction performance as illustrated in Figure 3.

The main limitation of our work is the use of virtual bodies with rigid body pose, uniform lighting, uniform texture, and the lack of spatial context. To generalize our findings to realistic cases, these limitations have to be addressed by including photos of children in a realistic setting.

4 Conclusions and Future work

In this paper we showed that estimating height and waist circumference from 2D projections of body scan images with ResNet-50 and Inception V3 is possible. From our results we conclude that automatic malnutration detection of single images seems feasible, provided one or more suitable viewpoints are used.

In this study we focused on predicting height and waist circumference in adults, as these metrics are represented by two identifiable PC values. When detecting malnutrition in children other body-shape characteristics, such as weight, upper-middle arm circumference and head circumference are more relevant. In order to include these metrics and tailor the methods further towards detecting malnutrition in children, in our future work we aim at expanding our measures, improve the realism of our dataset, and explore the usefulness of a shape prior. The aim of this paper is to propose an automated solution for a problem that is life threatening for millions of children around the world. Our call is not for an AI tool to replace human medical intervention, it is merely to have a remote tool for accessing those in hard to reach areas and to be used in times like Covid 19 where human interaction is limited.

Acknowledgments. This research is part of a research collaboration between Tilburg University's Zero Hunger Lab and Welthungerhilfe. Our research is inspired by Welthungerhilfe's Child Growth Monitor project that aims to develop a mobile application to measure body dimensions of children based on a single photo or video of the child. We thank the anonymous reviewers for their constructive comments on an earlier version of this paper.

References

[1] Local Burden of Disease Child Growth Failure Collaborators. Mapping child growth failure across low-and middle-income countries. Nature. 2020;577(7789):231.

[2] Holla R. The malnutrition bazaar: the case of RUTF. World Nutrition. 2021;12(2):104–118.

[3] Mertens E, Peñalvo JL. The burden of malnutrition and fatal COVID-19: a global burden of disease analysis. Frontiers in Nutrition. 2021;7:351.

[4] Li Z. 3D Human Pose and Shape Estimation Based on Parametric Model and Deep Learning. Lund University; 2021.

[5] Bogo F, Kanazawa A, Lassner C, Gehler P, Romero J, Black MJ. Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image. In: Computer Vision – ECCV 2016. Lecture Notes in Computer Science. Springer International Publishing; 2016. .

[6] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–778.

[7] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 2818–2826.

[8] Carletti M, Cristani M, Cavedon V, Milanese C, Zancanaro C, Giachetti A. Estimating body fat from depth images: Hand-crafted features vs convolutional neural networks. In: Conference and Exhibition on 3D Body Scanning and Processing Technologies; 2018. .

[9] Lee DS, Kim JS, Jeong SC, Kwon SK. Human height estimation by color deep learning and depth 3D conversion. Applied Sciences. 2020;10(16):5531.

[10] Zheng Z, Yu T, Wei Y, Dai Q, Liu Y. DeepHuman: 3D Human Reconstruction from a Single Image. In: The IEEE International Conference on Computer Vision; 2019. p. 7739–7749.

[11] Dantcheva A, Bremond F, Bilinski P. Show me your face and I will tell you your height, weight and body mass index. In: 2018 24th International Conference on Pattern Recognition (ICPR). IEEE; 2018. p. 3555–3560.

[12] Dibra E, Jain H, Öztireli C, Ziegler R, Gross M. Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks. In: 2016 fourth international conference on 3D vision (3DV). IEEE; 2016. p. 108–117.

[13] Yang Y, Yu Y, Zhou Y, Du S, Davis J, Yang R. Semantic parametric reshaping of human body models. In: 2014 2nd International Conference on 3D Vision. vol. 2. IEEE; 2014. p. 41–48.

[14] Pishchulin L, Wuhrer S, Helten T, Theobalt C, Schiele B. Building statistical shape spaces for 3d human modeling. Pattern Recognition. 2017;67:276–286.

[15] Haritosh A, Gupta A, Chahal ES, Misra A, Chandra S. A novel method to estimate Height, Weight and Body Mass Index from face images. In: 2019 Twelfth International Conference on Contemporary Computing (IC3). IEEE; 2019. p. 1–6.

[16] Li S, Nguyen VH, Ma M, Jin CB, Do TD, Kim H. A simplified nonlinear regression method for human height estimation in video surveillance. EURASIP Journal on Image and Video Processing. 2015;2015(1):32.