Jori Väinölä

VISION-BASED BODY MEASUMENTS

Optimization of prediction accuracy

Baschelor's Thesis
Faculty of Information Technology and Communication Sciences
February 2021

ABSTRACT

Jori Väinölä: Vision-based body measuments
Baschelor's Thesis
Tampere University
Computing and Electrical Engineering
February 2021

Recovery of 3D body information from 2D images is important yet challenging task with many applications in industrial design, clothing and online shopping and medical diagnosis. The development in this field is however increasingly slowed down by lack of publicly available data-set of realistic 3D human models. In order to avoid this issue, silhouette models are used, which are generated with a parametric body model learned from real 3D body scans.

The purpose of this study to evaluate which parameter influence the result most in predicting body measurements from 2D images. The tested parameters were the resolution of the images, the amount of the generated silhouette models, type of neural network, direction of image in relation to the body, and number of images per silhouette model. By studying these factors, the goal is to find the optimal combination of the parameters in order to best optimize the accuracy of the prediction model.

Results indicate that accuracy is highly correlated with training data size and functions best in combination with a simple convolutional neural network. Quality of the images, image angle, and use of multiple images also provide benefits, but affect is smaller compared to network and data size.

Keywords: Body measurement, prediction accuracy, parameter optimization

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

2D	two dimensional
3D	three dimensional
CNN	Convolutional neural network
PCA	Principal component analysis
RGB	Red green blue

1. INTRODUCTION

The recovery of 3D body information from human bodies is important yet challenging with many applications in industrial design [1], clothing and online shopping [2] and medical diagnosis [3]. This task can be achieved with 3D body scanning, but this technology is limited in cost, maintenance, mobility, and accuracy. However, this same task can be accomplished with 2D images and analysis of the body shape configurations.

The development in this field is however increasingly slowed down by lack of publicly available data-set of realistic 3D human models [4]. Models are used for gathering detailed information about body formations. The lack of realistic models can be solved using synthetic samples that are generated based on the distribution of measurements on real human models. This data is not as accurate as of the real 3D scans, but it does allow the use of a much larger sample size.

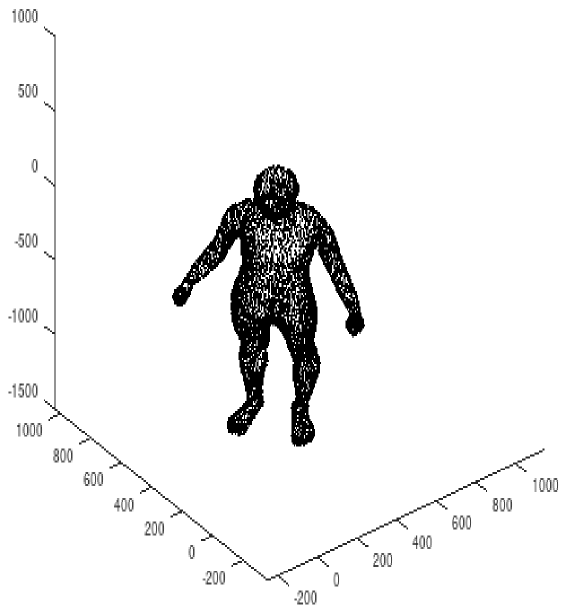

Figure 1.1. *Generated model represented in 3D space*

The purpose of this thesis is to study different factors that affect the accuracy of the body measurement prediction network. The network used in this study to make the predictions is based on a framework created by Johan Wirta, Song Yan, and Joni Kämäräinen [4]. The network operates by taking images of a human as input and produces estimates for different body part measurements.

The goal of this study is to identify what factors affect the accuracy of the prediction measurements and how they can be modified in order to optimize the accuracy. The different factors that are tested are the amount of the training images, resolution of the training images, type of neural network, angle of the images, and the effect of multiple images per one model. These parameters are tested with different values and evaluated how they effect one another and on what value range they work most optimally.

The study is structured as follows: Chapter 2 contains a brief mention of related work, Chapter 3 will explain the procedure being obtaining information from 2D images, chapter 4 will entail the tests that are conducted and how they are being evaluated, Chapter 5 presents the results for the tests and analysis of the result, and then final chapter gathers the main findings of the study for a conclusion.

2. RELATED WORK

There are multiple studies that also share the topic of image based detection of body features. These studies use different approaches like, generating a outline path from a silhoette image to identify key feature points [5], recognition of feature points from silhoette with mathematical formula [6], and combination of different methods for different feature points [7]. These studies also tackle the issue of creating silhoette from a real human image, which is not discusses in this study.

There are also other applications that feature similar goals to the ones of this study. One other application is the recognition of body posture from 2D images [8][9]. There are many studies on this topic and it is more commonly researched topic. Whilst sharing many similarities with the goal of this study, the recovery of shape and pose of a human does have their own challenges.

Similar technology is also used in animal husbandry. where machine vision is used to predict the growth of the animals. One of these application is prediction of weights of fishes and carcass traits. [10]. The application is based around similar idea of gathering information from silhouettes to form a prediction.

While the use of 3D technology might currently be limited by the unavailability and high cost of 3D scanning, this issue might be changed in the future. In the study from National Yunlin University of Science and Technology, a concept of creating a 3D recreation of an object with a single CCD camera is explored [11].It operates by mapping the layout from multiple images with machine learning based surface reconstruction. Through these methods it is possible to create a rapid low cost scanning system. These devices would would make 3D applications more accessable and potentially bring 3D technology to a more everyday use.

3. METHODS

Machine vision-based measurement prediction is built around a single purpose. Gathering information from a 2D image in order to make physical predictions on a human being. In this study, a convolutional neural network (CNN) is used to create a learning model for this purpose. CNN is a deep learning algorithm that be can used to identify important aspects from an image based on learning weights and biases [12]. The details of the implementation are presented in chapter 3.2.

The primary task of CNN is to identify body measurements from the image. 24 measurements to be exact. These include circumference and lengths of various body parts, with separate results for the left and right side of the body. The network identifies the segments representing different body parts from the image as seen in the Figure 3.1 and gives an estimation of their lengths in millimeters

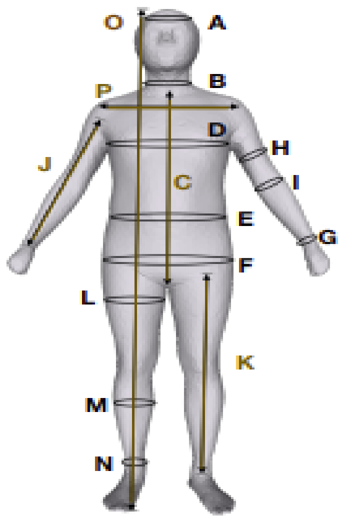

Figure 3.1. *Measurement points of the model [4]*

3.1 Data generation

The dataset used in this thesis is the CAESAR-dataset. CAESAR is a commercial sold dataset which contains thousand of measured male and female subject [13]. This dataset however cannot be used for public evaluation, because of a conmmercial license. So instead, a dataset created by Max Planck Institut is used. Models in this dataset are created from statictical body representation, learned from the CAESAR dataset [14]. Image of a model from the dataset can be seen at Figure 1.1.

The initial step for creating the prediction model is to split the data into two sets. The first set will be used as training material for the CNN for the measurement prediction. The second set will be used in testing to evaluate how accurate the prediction of the network are.

After the split there is training set of an 2200 models with each one having around 6500 3-dimensional points in a matrix. This is not enough for extensive testing and more must be generated. Generating is done by using GNU Octave, which is a software, primarely intended for numerical computations. On every model, each point represent roughly the same part of the body. Octave is used to read the data and perform centering on the matrixes. Afterwards, principal component analysis (PCA) is computated for the centered data, creating a new set of data with lower dimensions [15]. Afterwards, gaussian distribution is used to generate new random point in the PCA's range. With this, new models with random variation in their measurements are created.

The next step is to calculate the measurements from the newly created models. This is accomplished by measuring distances between points for lengths and paths on the models surface for circumferences. In this study, measurements are acquired through pre-marked arrangements of points. This will results in a dataset of measurements that will be used to train the CNN.

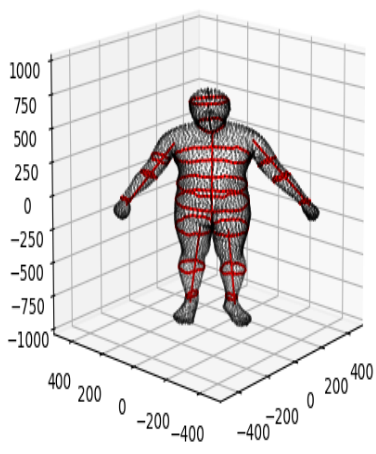

Figure 3.2. *Measuring of the models features*

The last piece of required data are the matching images for the dataset. The images used are silhoettes, meaning they are a representation of an object uniformally filled in one color, outlining it from the background. This is used to train the network to recognize models features from its outline. Silhoettes are instead of normal RGB images, because of lack of a proper dataset. The images are created by first rendering the models with Python and then generating a drawing of the model from specified angle, as seen in the Figure 3.3.

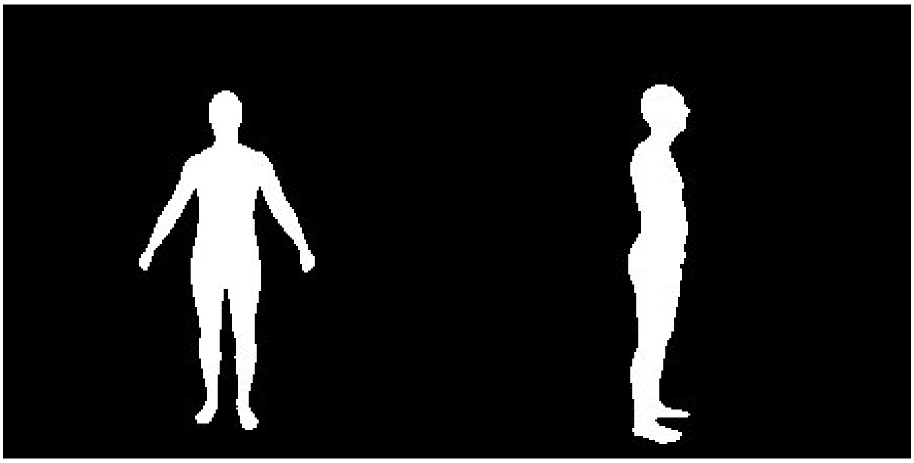

Figure 3.3. *Silhoette images generated from the 3D model*

3.2 Network

With the methods of chapter 2.1, the required data for the vision based body measurement has been received. The next step in process is to create a learning model which will use the data and learn how to predict measurements from a 2D image. In this study, this is done with a convolutional neural network. This CNN is implemented with Keras, which is a python library.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 16)	160
max_pooling2d (MaxPooling2D)	(None, 112, 112, 16)	0
conv2d_1 (Conv2D)	(None, 112, 112, 32)	4640
max_pooling2d_1 (MaxPooling2	(None, 56, 56, 32)	0
conv2d_2 (Conv2D)	(None, 56, 56, 32)	9248
max_pooling2d_2 (MaxPooling2	(None, 28, 28, 32)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 32)	802848
dense_1 (Dense)	(None, 24)	792

Figure 3.4. *Layers of the convolutional neural network*

The layers contained in the base network are presented in the Figure 3.4. Network is modified in the test, but it always contains the same operations, but their amount will change. The conv2d layers create a convolution that convolved with the layers inpurs to produce a tensor of outputs, maxpooling downnsizes the inputs by taking max values over a window for each dimension, flatten reshapes the input into one dimension, and dense is a regular deeply connect neural network layer [16]. The structure is trained with combination of silhouette images and measurements to produce measurement prediction from image inputs.

After the training, the final step is to test the network. This is done using silhoettes of the models that were originally assigned for testing. The image are used as the input for the network and the outcome is compared to calculated values from the 3D models.

Measurement name	Avg value (mm)	Mean error (mm)
A head circ	603.04	12.47
B neck circ	393.62	11.25
C crotch len	708.45	9.38
D chest circ	1077.98	49.29
E waist circ	957.63	54.99
F pelvis circ	1031.4	32.32
G wrist circ left	181.61	10.3
G wrist circ right	185.4	9.77
H bicep circ left	321.82	14.64
H bicep circ right	315.82	12.9
I forearm circ left	288.51	13.99
I forearm circ right	327.3	12.8
J arm len left	530.34	9.36
J arm len right	538.77	10.25
K leg len left	704.87	25.05
K leg len right	709.45	12.64
L thigh circ left	614.63	27.59
L thigh circ right	605.39	20.89
M calf circ left	415.68	9.62
M calf circ right	403.84	11.11
N ankle circ left	247.79	8.6
N ankle circ right	234.44	7.23
O overall height	1767.38	25.65
P shoulder breadth	403.66	8.3

Table 3.1. *Example output of the network containing the average estimate for body segments and mean error for them*

4. EXPERIMENTS

The goal that this study is to find what are the most important design choices for the network in terms of improving the prediction accuracy. The tested parameters are resolution, training data size, angle of the body, architecture of the network, and effect of multiple images. The primary goal of the testing is to find if modifying parameters has a correlation with the prediction accuracy or if there are any ranges where specific parameters provide exceptional results.

Testing is conducted through few cycles of individual and paired testing. If the testing were done by measuring every parameters affect on others, the experiment durations would grow too large for capabilities of this study. For that reason, the amount of tests must be kept limited. To limit the amount of combination that will need to be run in test, the testing will be conducted with following method in mind. On every cycle one or two parameters will be tested. Results of the round are analysed and the best performing values will be used for the later rounds. After all the cycles, resulting combination should be improved in every factor from the original parameters.

Chapter 3.1 will discuss the evaluation process of the experiments, and Chapters 3.2, 3.3, and 3.4 will explain the factors being tested in more thorough detail.

4.1 Performance evaluation

As seen in the table 3.1, the outcome provides multiple values for error. In order to compare results, a value is need to measure the prediction accuracy. This value could be a selection one or multiple measurements from the results, but this might put too large emphasis on the body parts with largest measurements. In order to avoid this issue, a percentage error is going to be used instead of absolute values. The error percentage is calculated from mean error divided by average value on each body part. Since it is also unclear what body parts wary the most and which represent the overall quality the best, the error is measured with combined value of average error in all the body parts. The formula can be seen in the Equation 4.1. The lower the value is, the less there is error and thereby the prediction is more accurate. The formulas use in practice is shown below in the Table 4.1.

$$Total\_percentage\_error = \sum \frac{mean\_error_i}{avg\_value_i} \tag{4.1}$$

Measurement name	Avg value (mm)	Mean error (mm)	percentage error
A head circ	603.04	12.47	2.07
B neck circ	393.62	11.25	2.86
C crotch len	708.45	9.38	1.32
D chest circ	1077.98	49.29	4.57
E waist circ	957.63	54.99	5.74
F pelvis circ	1031.4	32.32	3.13
G wrist circ left	181.61	10.3	5.67
G wrist circ right	185.4	9.77	5.27
H bicep circ left	321.82	14.64	4.55
H bicep circ right	315.82	12.9	4.08
I forearm circ left	288.51	13.99	4.85
I forearm circ right	327.3	12.8	3.91
J arm len left	530.34	9.36	1.76
J arm len right	538.77	10.25	1.9
K leg len left	704.87	25.05	3.55
K leg len right	709.45	12.64	1.78
L thigh circ left	614.63	27.59	4.49
L thigh circ right	605.39	20.89	3.45
M calf circ left	415.68	9.62	2.31
M calf circ right	403.84	11.11	2.75
N ankle circ left	247.79	8.6	3.47
N ankle circ right	234.44	7.23	3.08
O overall height	1767.38	25.65	1.45
P shoulder breadth	403.66	8.3	2.06
Total		420.39	80.1

Table 4.1. *Example output of the network containing the average estimate, mean error, and percentage error for each bodypart*

The initial setting in training is 10000 images with a resolution of 220x220. The initial network will be the one shown in the Figure 3.4 . The results of this test is used as a baseline to measure how large the overall improvement in this study is.

4.2 Image resolution

The first factor that is tested is the resolution of the image. By testing images with different resolutions, it should be possible to determine if amount of information retrieved from the image is highly dependant on the pixel size. With this it should also be possible to see at what point do details become so small that they do not provide more information.

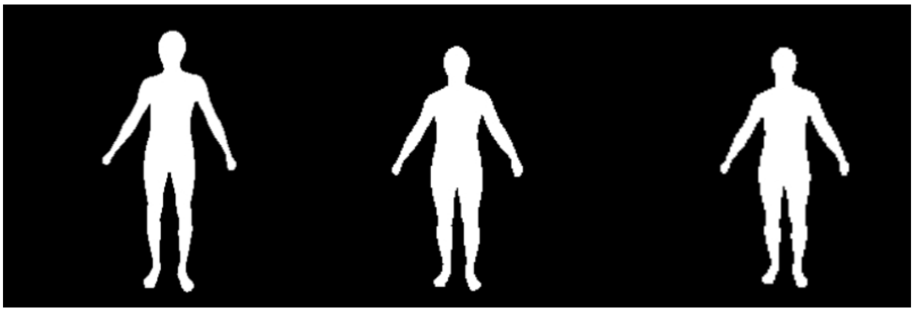

Figure 4.1. *Silhouette images with resolutions of 300x300, 240x240, and 180x180*

Image quality is the first parameter being tested and therefore it is possible that prediction accuracy improvements from changes in resolution are also dependant on other parameters. Because of this resolution tests are conducted using multiple different network architectures.

4.3 Number of training images and network architecture

Second factor that will be tested is the amount of data used to train the neural network. The goal is to find how much data is needed to sufficiently train the network and after what point does the training start providing diminishing return. The intent is not to find a specific amount, but a level at which these points start to occur. For this reason, the test are started at 10000 images and the patch size grown exponentially every cycle.

The required amount of training is also highly reliant on the complexity of the used network. Because every additional layer in the neural network means more parameters that have to be optimized. these additional parameters to be efficiently trained, an additional training data may be required. Because of all this, data quantity and the network complexity are tested together

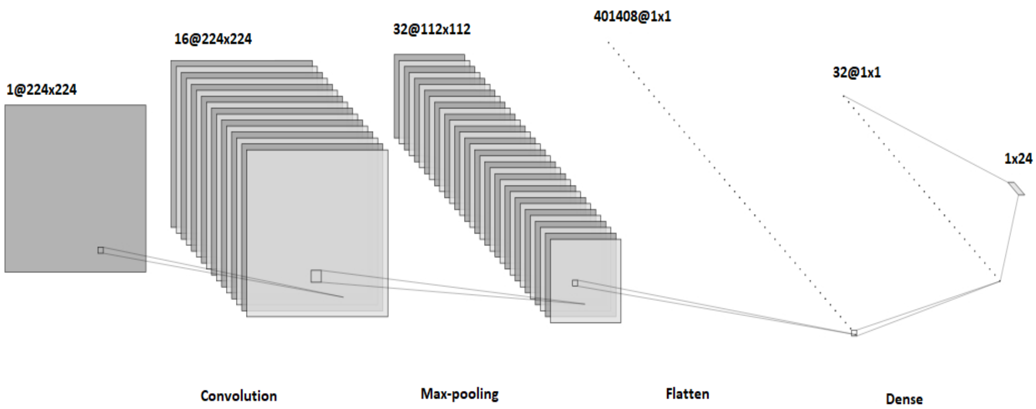

Figure 4.2. *Network structure with one convolutional layer*

The network complexity is tested by adding and decreasing the amount of layers in the structure. The layers that are implemented in these operations are Conv2d and max-pooling2d. The tested network will contain N amount of these layers with N varying each experiment.

4.4 Body angle

Angle of the images might also provide differences in accuracy since some body details are better observed from certain angles. Angle is tested by taking images every 15 degrees over a 180 degree arc from models front to back. Test only cover half a circle because tests are done with the assumption that left and right side of the human body are identical and do not require separate testing. The angles are better shown in the following Figure 4.3.

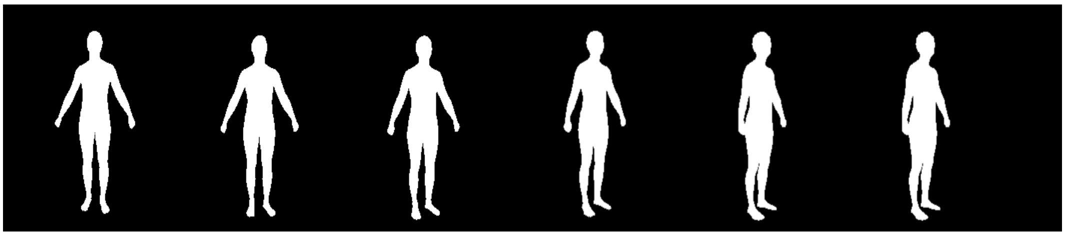

Figure 4.3. *Silhoettes from different angles*

After the best angle has been found, the next step is to study what is the best combination when two images are as input instead of one. This is simply done by testing all the other angle combination with the best performing on from the first experiment.

Use of multiples input does however pose some difficulties. This increase of input demands changes to the network that can accomodate the large size of data. Due to these reasons, the network structure has to be changed for the final tests. The new network will accept two separate input that are run through some convolutions. After this the results will be combined using Keras' maxinum function and passed through another set of convolutions.

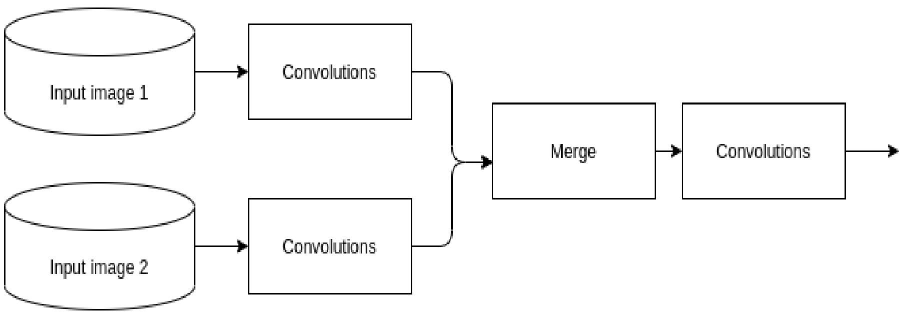

Figure 4.4. *Network structure with 2 input images*

With the change of structure the optimal network complexity may not match to the one in the earlier experiments. Because of this, the complexity of first and second part of layers will be adjusted according to earlier experiments. The results of the two image network will be compared to the network with one input.

5. RESULTS

In this section we report the results from the various experiments. Some of experiments were modified and some were performed multiple times according to the results. The order of the tests is still the same as mentioned in the chapter 4.

5.1 image resolution

The first experiment was to test how quality of the image effects the result and find logical explanations for the phenomenon. The images that are tested range from 180x180 to 340x340 pixels, with test every increment of 20. These test are conducted with different network ranging from 3 to 5 convolutional layers.

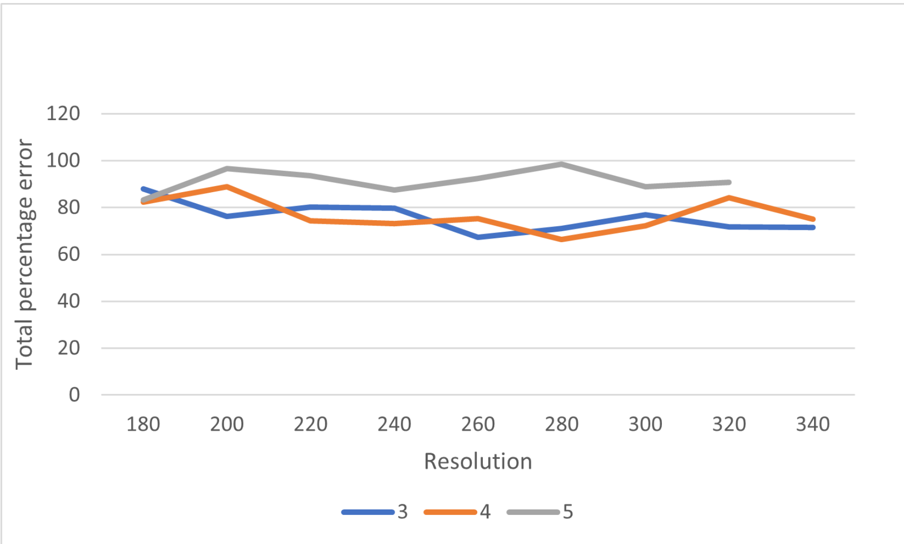

Figure 5.1. *Total percentage error from experimenting different resolutions with three network architectures*

The data appears to fluctuate which makes it difficult to draw conclusions. With 3 and 4 four layers it seems that the most accurate results appear in the 260 to 280 range. With 5 layers, the accuracy gets worse when increasing the resolution. To make more certain assumptions, another test is conducted with similar settings with exception of larger training set and more layered network.

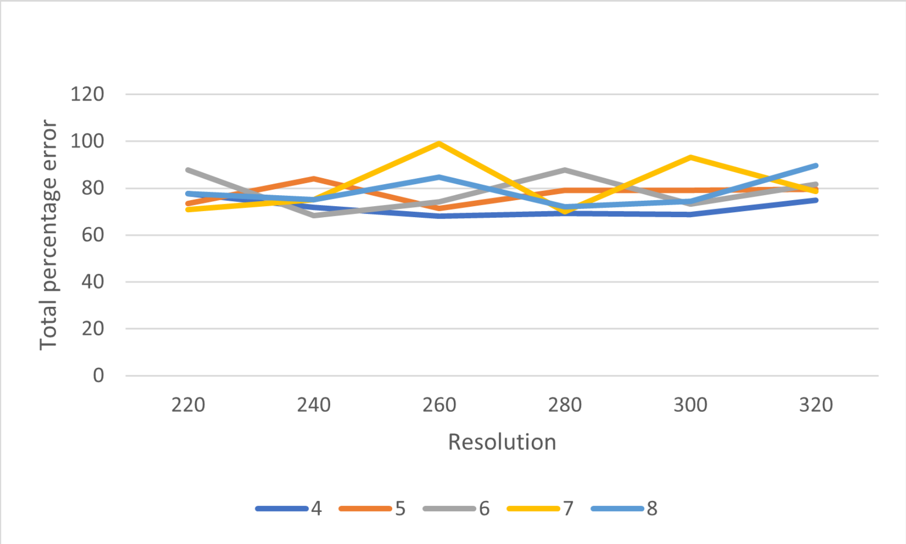

Figure 5.2. *Repeating image resolution experiments with training data size of 20000*

According to the new results shown in the figure 5.2, it appears that change in accuracy seems to stagnate 280 to 300 resolution range on networks with lower amount of layers. Most network do still hit their peak in the 260 to 280 range. Since most network reach their best accuracy at 280 and with the sign of stagnation when moving to higher sizes, 280 is picked as the standard image size for the later tests.

Based on the limited testing, there did not appear to be a clear linear behavior in image resolution and the prediction accuracy. Because of this, the only information used from the results is the point at which most network appear to reach their highest accuracy. The selected size of 280 pixels corresponds one pixel being 7.5mm x 7.5mm.

5.2 network architecture and training data size

The second experiment contains testing of training data quantity together with networks of different complexity. Goal here is to find how much does increase of training data improve the quality and see what kind of network appears to provide to most optimal results. The values that a used for the training data are 10k, 20k, 40k, and 80k. This is paired with network containing 1 to 5 convolution layers.

Figure 5.3. *Testing training data sizes with different network architectures*

Figure 6.3 shows how the number of images effect accuracy. Increasing training data provides improvement to the network an all complexities. The amount of images used was grown exponentially in this test, so to take this effect further would require a large amount of data. The effect of increasing training data has a larger effect when used on more complex network, but these networks appear to hit their limit at 40000 images. After this point data size provides very little benefit, but to the more simpler network there still appears to be a small but significant gain from using 80000 images.

Second point that can be gathered from Figure 5.3 is how the different network compare against one another. The less complex networks generally provide greater results and in cases where this is not true, the difference between network is small. Effect between networks gets smaller with larger training, but with the highest training data size the smallest architecture still provides the best result. This does bring up a new question. To what extend can the network be simplified until a point is reached where accuracy start to worsen again. The one layer convolution layer network can be simplified even more by reducing the amount of filters inside it. The test this far have been using layers with 32 filters each. At same time we are testing using kernel sizes of 5 and 3.

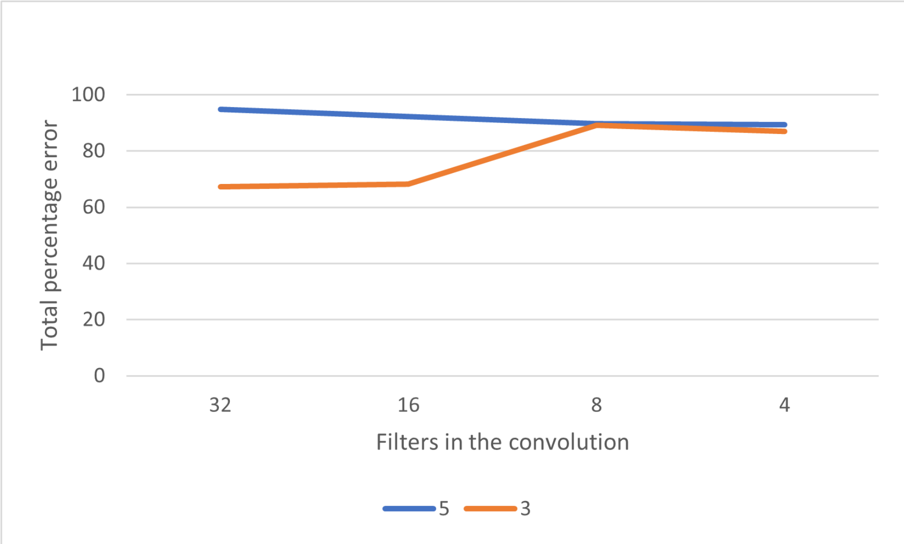

Figure 5.4. *Testing filter sizes with one convolution layer*

The kernel size of 3 clearly provides more optimal results. Both kernal sizes accuracy behaves relatively the same on small amount of filters but the clear winner is 32 filters. The difference between 32 and 16 layers is very small. Based on the figure 5.4, the one layered network from the earlier figure 5.3 is also the most optimal.

5.3 body angles and multiple images

In this experiment we verify whether the angle of the body has large effect on prediction accuracy. Testing is simplified with the assumptions that the models left and right side mirror each other and therefore only experimenting in an arc of 180 degrees for one side are necessary. Images are taken every 15 degrees with 0 degrees meaning straight from front.

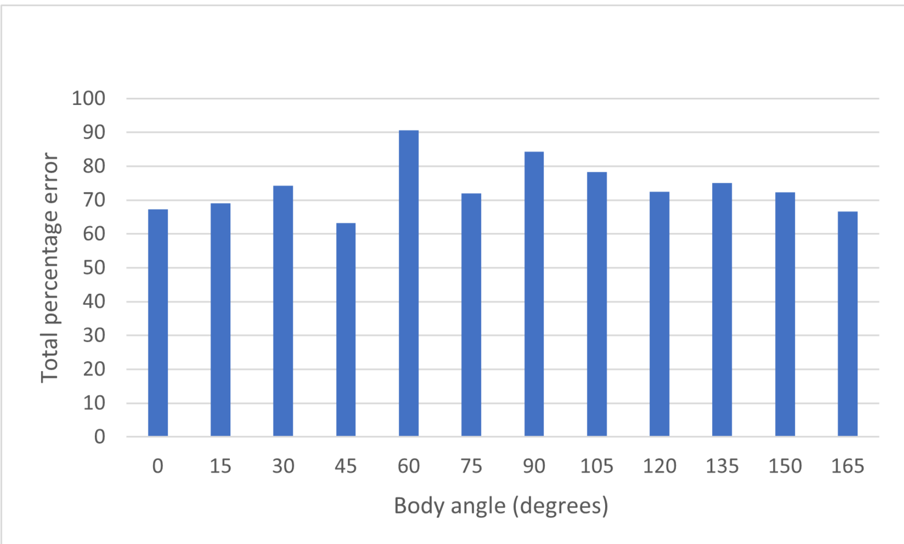

Figure 5.5. *Experimenting with different body angles*

The test shows that there is a large difference between angles and their accuracy. Images from either straight from front or back plus a small angle provide the best results, with the exception of 45 degrees which is the most accurate. Some of the difference between angles might be caused by the fact that in silhouette images some body features might blends into one another and therefore are harder to recognize from certain angles.

Next experiment will include the use of multiple images, with goal of finding the best combination of angles. Based on the earlier results, network will only use minimal amount of convolution layers. The network will consist of one convolution layer before and after the merging of input data. The first images angle is 45 degrees, which is the best angle from the figure 5.5. The use of two 45 degrees is skipped.

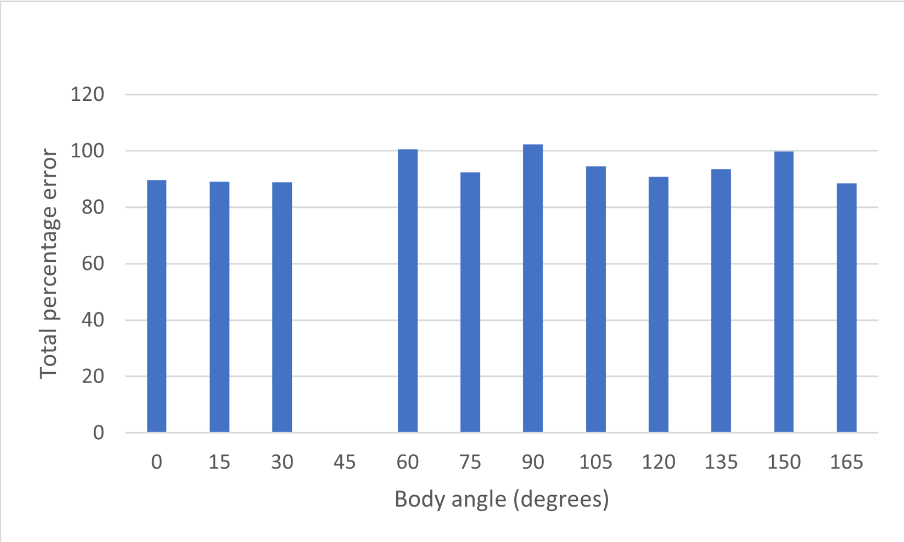

Figure 5.6. *Experimenting with second images body angle*

According to the test, the best combination of angles is 45 and 165. The difference between angles is much shorter than in the Figure 5.5. This gives an indication that angle of the second image is not as important compared to selecting angle for one image network. From the figure, it can also be noted that accuracy has gone down from just using one image. This could be caused by lack of training if network requirement change when doubling inputs. This is tested by experimenting the best combination of angles with a higher training data size. The results is compared to optimal combination for one image to see if multiple angle provides advantages over it.

Figure 5.7. *Comparing architectures with one and two input images.*

Two image network gets a large benefit from increase of training data. As seen in the figure 5.7, two image network has much lower accuracy at 10000 images, but rapidly improves with more images and at 80000 it bypasses one image entirely. Based on the improvement of two image network with training, it would be likely that further increase of training data would improve the accuracy even further.

6. CONCLUSION

In this study, image based measurement prediction was experimented using various combinations of parameters. The parameters that were tested were the amount of training data, quality and angle of the images, effect of multiple body angles, and network architectures. Experiments were performed with different combination of values and the results were evaluated based on combined percentage error of different body measurements. The results were presented visually.

First test evaluated the effect of different pixel sizes for the images. Image sizes were tested in combination with different training data sizes and multiple networks. Results show that the pixel size does impact the accuracy, but there does not seem to be any correlation between growth of image size and effect on end result.

Second test was the combination of training data sizes with networks of different complexities. Growing the amount of training data seemed to have positive effect on all levels, but this effect seem to provide smaller benefits when data size is continously raised. The architectures all showed similar behavior with the training data increase, but the best performing network seem to be smallest, with only one convolution layer proving to be the most effective.

The third experiment included the use of different angles and combination of multiple images. When using only one image, the effect of the angle seem to be quite large. The best angle was at 45 degrees from front, but images from front and back with slight deviations also provided as nearly accurate results. When using multiple images, the angle of the second image had only very minor differences on the accuracy. Using two images did however have a decrease in the accuracy overall, but this effect could be countered by using larger training data size. With enough training, two image model proved more accurate then one image model.

The absolute best combination of parameters found from the test would be two 280x280 images with angles 45 and 165, with only convolution layered network and training data of 80000 images.

measurement name	Avg error (original)	Avg (optimized)
A head circ	12.47	8.6
B neck circ	11.25	8.63
C crotch len	9.38	8.56
D chest circ	49.29	18.68
E waist circ	54.99	17.87
F pelvis circ	32.32	10.97
G wrist circ left	10.30	8.83
G wrist circ right	9.77	7.46
H bicep circ left	14.64	10.65
H bicep circ right	12.9	11.11
I forearm circ left	13.99	9.55
I forearm circ right	12.8	8.22
J arm len left	9.36	4.70
J arm len right	10.25	6.54
K leg len left	25.05	8.18
K leg len right	12.64	6.97
L thigh circ left	27.59	11.43
L thigh circ right	20.89	9.21
M calf circ left	9.62	6.84
M calf circ right	11.11	6.87
N ankle circ left	8.6	6.21
N ankle circ right	7.23	5.98
O overall height	25.65	10.75
P shoulder breadth	8.3	7.24
total	420.39	220.07

Table 6.1. *Average errors (mm) from network run with the original parameters and one with the optimized parameteres*

The overall improvement when comparing the initial state and the optimized state, shows a promising reduction of 48% in the average error. This with the combination of findings in parameters effect stands to prove that this study has reached its goal. The study was limited to quite small testing sizes, but the great overall improvement in the prediction accuracy would only seem to indicate that with a wider testing, the optimization of parameters could be improved to an even further degree.

REFERENCES

[1] B.-K. D. Park, S. Ebert, and M. Reed. "A parametric model of child body shape in seated postures", *Traffic Infury Prevention* ,vol.18, no.5 (2017).

[2] H. Daanen and S.-A. Hong. "A parametric model of child body shape in seated postures". *International Journal of Clothing Science and Technology* ,vol.20, no.1 (2008).

[3] C. Ogden, C. Fryar, M. Carroll, and K. Flegal. "Mean body weight, height, and body mass index, united states 1960–2002", *Division of Health and Nutrition* (2004).

[4] Song Yan, Johan Wirta, Joni-Kristian Kämäräinen. "Silhouette Body Measurement Benchmarks". *In International Conference on Pattern Recognition (ICPR)* (2020).

[5] Yueh-Ling Lin, Mao-Jiun J. Wang. "Automated body feature extraction from 2D images". *Expert Systems with Applications* (2011).

[6] Ouellet, François Michaud. "Enhanced automated body feature extraction from a 2D image using anthropomorphic measures for silhouette analysis". *ScienceDirect* (2018).

[7] Muditha Senanayake, Amar Raheja and Yuhan Zhang. "Automated human body measurement extraction". *International Journal of Clothing* ,vol.30, no.2 (2018).

[8] Kanazawa, B. “End-to-end recovery of human shape and pose". *CVPR* (2018).

[9] J. Tan, I. B. and Cipolla, R. “Indirect deep structured learning for 3d human body shape and pose prediction". *BMVC* ,vol.3, no.5 (2016).

[10] Fernandes, T. “Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia". *Computers and electronics in agriculture* (2020).

[11] Yaan Zhang, Zhankun Luo, Jintao Hou, Lizhe Tan, Xinnian Guo. "Computer Vision Techniques for Improving Structured Light Vision Systems", *Electro Information Technology (EIT) 2020 IEEE International Conference on* ,pp. 437-442 (2020).

[12] Sumit Saha. *"A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way"*. URL: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[13] *"The most comprehensive source for body measurement data"*. URL: http://store.sae.org/caesar/ (visited on 12/06/2020).

[14] Pishchulin, L., Wuhrer, S., Helten, T., Theobalt, C. and Schiele, B. Building Statistical Shape Spaces for 3D Human Modeling. *Pattern Recognition* (2017).

[15] Jaadi, Z. *"a step by step explanation of principal component analysis"*. URL: `https://builtin.com/data-science/step-step-explanation-principal-component-analysis` (visited on 12/06/2020).

[16] *"Keras layers API"*. URL: `https://keras.io/api/layers/` (visited on 12/06/2020).