

# Non-Contact Human Body Dimension Estimation Methods Based on Deep Learning: A Critical Analysis

**Yichen Zhang**

School of Artificial Intelligence, South China Normal University, Guangzhou, China

20212005177@m.scnu.edu.cn

**Abstract.** With the advancement of information technology, Human Body Dimension Estimation (HBDE) is moving towards digitalization and non-contact methods. This paper critically reviews non-contact HBDE methods based on deep learning, primarily covering image-based methods, the Adaptive Body Structure Segmentation (ABSS) algorithm with height and weight measurements, and 3D meshes or point cloud data methods. Convolutional Neural Networks (CNNs) excel in image measurement, while the ABSS algorithm with adaptive segmentation achieves high accuracy. Recent studies, such as the Grey Wolf Optimizer-Elman Neural Network (GBWO-ENN) model, have significantly improved prediction accuracy by optimizing neural network structures. Additionally, 3D point cloud data combined with the Skinned Multi-Person Linear (SMPL) model has proven effective in complex scenarios. This paper summarizes the performance of various methods, analyzes common datasets, and explores future development directions for non-contact HBDE. Despite significant progress, this field still faces challenges such as data quality, scarcity of extreme samples, and privacy issues. Future research should focus on building high-quality datasets and developing lightweight measurement networks to meet practical application needs.

**Keywords:** Non-contact measurement, human body dimension estimation, deep learning, convolutional neural networks.

## 1. Introduction

With the rapid development of information technology, Human Body Dimension Estimation (HBDE) is advancing towards digitalization and non-contact methods. Human body dimensions refer to measurements of various parts of the body such as height, weight, or more detailed metrics like shoulder width or chest circumference. Reliable methods for measuring the former, such as scales for weight, already exist. However, methods for measuring the latter often have significant errors in everyday scenarios despite being of greater reference value in primary applications such as clothing design.

Currently, non-contact HBDE methods mainly rely on 3D body scanning devices[1] or combine human depth data captured through RGB-D cameras. However, these methods face challenges such as high costs, stringent environmental requirements, and poor portability. Research on non-contact HBDE is still relatively scarce, and comprehensive reviews are particularly lacking. Therefore, this paper aims to provide a comprehensive, extensive, and in-depth summary of non-contact HBDE methods based on deep learning.

Starting from image-based measurement methods, methods based on the Adaptive Body Structure Segmentation (ABSS) algorithm with height, weight, and circumference values, and methods based on 3D meshes or 3D point cloud data, this paper reviews recent research achievements in the field of non-contact HBDE. It analyzes and compares the main models and performances of various methods and looks forward to the future development of HBDE.

## 2. Related Works

Non-contact HBDE relies on neural networks trained on a large number of human body dimension measurement data samples. This section introduces some key related works. Table 1 summarizes several models used for HBDE and their backbone networks.

**Table 1.** Network models and its backbone

| Network Models       | Backbone     |
|----------------------|--------------|
| Neural Anthropometer | CNN          |
| Shapy                | CNN          |
| BMnet                | CNN+GBWO-ENN |
| GBWO-ENN             | ENN          |
| IWOA-ENN-MC          | ENN          |
| UGA-BP-MC            | BP           |
| PC-BoDiEs            | MLP          |
| Conv-BoDiEs          | CNN          |

Human Body Dimension Estimation: Non-contact methods for HBDE analyze images taken from different perspectives to measure various body dimensions without direct contact with the subject. In 2016, Dibra et al.[2] first attempted to use Convolutional Neural Networks (CNNs) to recover 3D human meshes and body measurements from silhouettes, training on synthetic silhouette images generated from the CAESAR dataset. In 2020, Hu et al.[3] proposed a human body multi-feature point extraction and dimension measurement algorithm based on ABSS, called Human pesm-abss, which accurately acquires dimension information of various body parts. In 2021, Hu et al.[4] introduced a model based on a BP network, integrating the Upgrade Genetic Algorithm(UGA) and Markov Chain(MC) method, resulting in a simpler neural network structure. In 2022, Nataniel Ruiz et al.[5] proposed using adversarial generative networks to increase the number of extreme samples and combine multi-view images, significantly reducing measurement errors. In 2024, Yang et al.[6] proposed a human body dimension prediction method based on an Elman Neural Network (ENN) network, which improved global stability and prediction accuracy compared to Hu's Upgrade Genetic Algorithm-Back Propagation-Markov Chain(UGA-BP-MC)[4] and the 2023 Improved Whale Optimization Algorithm-Elman Neural Network-Markov Chain(IWOA-ENN-MC)[7] neural network models.

Skinned Multi-Person Linear model(SMPL): SMPL[8] is a statistical model for 3D human shape and pose estimation based on 3D mesh vertices. It aims to capture the geometric shape and motion pose of the human body by training on large-scale 3D human scan data, and it is widely used in the field of computer vision.

Human pesm-abss algorithm (referred to as ABSS algorithm below): This is a method for extracting feature points and measuring human dimensions from 2D images based on Adaptive Body Structure Segmentation[3]. The method combines height proportion and image preprocessing techniques to address issues with traditional image detection models, such as long detection times and inaccurate key points. This method can obtain relatively accurate human height, gender, and circumference width and thickness results from 2D human images.

## 3. Summary of Non-Contact HBDE Methods

Generally, non-contact HBDE methods based on deep learning often rely on neural networks, with different methods employing different types of neural networks. Additionally, various non-contact

HBDE methods often require different types of input data, which has significant implications for the choice of methods in different application scenarios. Therefore, this paper categorizes these methods based on the types of input data required during the prediction process into image-based human dimension estimation methods, estimation methods based on the ABSS algorithm[3] with height, weight, and circumference values, and estimation methods based on 3D meshes or 3D point cloud data. This section will introduce representative methods from these categories, and the next section will compare their performance based on prediction results from specified datasets.

### 3.1. Image-Based Human Dimension Estimation Methods

Classic image-based measurement methods use CNNs to extract features from RGB images. These extracted features are further processed to generate 3D human meshes based on these parameters or detect key points using specially trained neural networks. Traditional methods may also involve manual operations to obtain key point positions of the human body, including shoulders, elbows, wrists, etc. The key points obtained by the above methods are finally used to infer specific human measurement data through geometric calculations and regression models. This subsection will briefly introduce two representative methods and describe the improvements they have made based on the above basic implementation logic.

*3.1.1. BMnet.* This method[4] uses Generative Adversarial Networks (GANs) and CNNs for human dimension estimation, enhancing model robustness through adversarial samples. First, preprocessed front and side human images are input, and a unique adversarial body simulator is introduced to fill the gap of extreme individual data. After feature extraction by CNN, the extracted feature maps generate high-dimensional feature vectors through multiple layers of convolution and pooling operations. Subsequently, the feature vectors are input into a key point detection network to detect the coordinates of human key points, including shoulders, elbows, wrists, hips, knees, and ankles, and further infer specific human measurement data. Finally, the model outputs highly accurate human measurement results, demonstrating the effectiveness of using adversarial samples and multi-stage key point regression methods in human dimension estimation. Additionally, the model also includes tests on multi-view inputs and height and weight inputs, all achieving ideal results.

*3.1.2. Shapy.* This method[9] uses CNNs to extract human features from RGB images, which are used to regress the parameters of the human model. Shapy still employs the method of generating 3D human meshes and recognizing human key points to obtain measurement values. Additionally, Shapy introduces a method to enhance model training through crowd-rated language shape attributes. By collecting linguistic descriptions of human shapes, such as "tall," "slim," "muscular," etc., and pairing them with the obtained 3D shapes, Shapy introduces loss functions related to language shape attributes in training, optimizing the model multidimensionally in combination with measurement value loss functions.

### 3.2. Estimation Methods for Height, Weight, and Circumference Values Based on the ABSS Algorithm

Compared to the image-based human dimension estimation methods mentioned in the previous section, methods based on the ABSS algorithm for height, weight, and circumference values usually do not use CNNs as their core neural networks. Unlike traditional methods, these methods do not involve generating or using 3D body models but instead perform adaptive segmentation of key areas of the human structure based on orthogonal human images to obtain gender, height and weight values, and circumference values, and directly predict the outputs. Additionally, some optimization algorithms or stability and fluctuation handling methods are also included to optimize the neural network. This subsection will mention several representative methods and briefly introduce the neural networks and optimization methods they employ.

**3.2.1. GBWO-ENN.** Grey Black Wolf Optimization–Elman Neural Network(GBWO-ENN) method[6] improves the traditional Grey Wolf Optimization Algorithm (GWO) by addressing its problems of easily falling into local optima and balancing global and local searches. The proposed GBWO algorithm uses a nonlinear decreasing method to reduce the convergence coefficient, thereby optimizing the algorithm's performance. GBWO is then used to optimize the ENN, using the weights and thresholds output by the trained GBWO model as the initial parameters for ENN training, achieving good accuracy and robustness in human dimension prediction.

**3.2.2. A-BP-MC.** This method[4] addresses the issues of the widely used BP neural network for human dimension prediction, such as easily falling into local optima and improper initial weight and threshold assignments. By optimizing the BP neural network with the improved UGA, and using the Markov Chain method to improve model accuracy and prediction stability in unstable environments. The main execution steps of the model include input image, body type classification, adaptive segmentation of key areas of the human structure based on the ABSS algorithm[3], feature dimension extraction, UGA-BP model training, and Markov correction model training, achieving good robustness and accuracy in human dimension prediction.

### **3.3. Human Dimension Estimation Methods Based on 3D Meshes or 3D Point Clouds**

Human dimension estimation methods based on 3D meshes or 3D point clouds often face challenges such as high data acquisition difficulty, significant cost, and scarcity of datasets. Alldieck et al. [10-12] proposed using depth cameras to acquire human depth information, then reconstructing the human body through the SMPL model and the Poisson method[13] to obtain circumference information of various parts. In the methods introduced in this section, 3D point clouds are directly used as input data, and the SMPL model is used to generate 3D human data and obtain measurements of various parts, which are then used for subsequent model training and validation. This subsection will introduce two typical human dimension estimation methods.

**3.3.1. Neural Anthropometer.** This model[14] proposes a classic human dimension estimation method using data generated by the SMPL model to obtain measurements of various parts of the human body through 3D modeling. Combined with the Blender software package, it synthesizes 2D images to train CNN, enabling the estimation of human dimensions from input 2D images. This method overcomes the scarcity of traditional human dimension data while providing a reliable human dimension estimation method.

**3.3.2. Point Cloud Body Dimensions Estimation (PC-BoDiEs).** This model[15] uses stacked Multi-Layer Perceptron (MLP) convolution layers to extract global and local features from point cloud data and regress 16 predefined human dimensions. Each MLP layer is followed by a ReLU activation function to enhance the model's ability to handle nonlinear problems. The input to the PC-BoDiEs model is unorganized 3D body scan data merged from two viewpoints. This data is sampled using farthest point sampling to match a fixed number of points before being fed into the network, reducing the model's time and memory requirements. During training, the model parameters are optimized using the AMSGrad variant of the Adam optimizer. Throughout the training process, the learning rate is gradually reduced using a cosine decay strategy. The PC-BoDiEs model, by processing data from unorganized 3D point clouds, overcomes the cumbersome process of fitting human scan data to predefined body models, directly regressing body dimensions from point cloud data, effectively improving the accuracy and efficiency of human dimension estimation from 3D data.

## 4. Major Datasets and Experimental Results Comparison

### 4.1. Major Datasets

Table 2 introduces commonly used datasets in HBDE tasks. According to the number of samples, the number of measurement items, and the types of data included in different datasets, they are suitable for various HBDE tasks.

**Table 2.** Commonly used datasets in HBDE tasks, including the datasets' name, number of samples, number of measurement items and the type of data used

| Dataset Name | Number of Samples (People): | Number of Measurement Items | Data Type                  |
|--------------|-----------------------------|-----------------------------|----------------------------|
| CAESAR       | 2400                        | 40+                         | 2D, 3D, Measurement Values |
| MAD          | 4419                        | 5                           | 2D, Measurement Values     |
| BodyM        | 2505                        | 14                          | 2D, 3D, Measurement Values |
| ANSUR II     | 13000+                      | 100+                        | Measurement Values         |
| Human3.6M    | 17                          | -                           | 2D, 3D                     |

- (1) CEASER (Civilian American and European Surface Anthropometry Resource): This dataset includes samples of adults from North America and Europe, providing detailed 3D body scans and body measurement data. It is widely used in generating 3D human mesh data using the SMPL model.
- (2) MMTS (Model Measurements Test Set): This dataset comes from several modeling agency websites and includes high-resolution multi-view human RGB images and detailed body measurement data. It was used in the training of the Shapy model [10], where researchers added semantic labels.
- (3) BodyM Dataset: This is an advanced dataset specifically for human pose estimation and 3D human reconstruction. It contains high-precision 3D body scans and pose annotations. The provided 3D human data includes 50 key points and was collected in a laboratory environment using different camera angles and lighting conditions to simulate outdoor image capture. Some subjects were photographed multiple times in different clothing to enhance the robustness of the dataset.
- (4) ANSUR II (Anthropometric Survey of U.S. Army Personnel): This dataset is a body measurement survey conducted by the U.S. military, aiming to collect and analyze the anthropometric data of U.S. soldiers. It has a large number of samples, containing over 100 body measurement items, but it does not include images, and all subjects are U.S. military personnel.
- (5) Human3.6M Dataset: This dataset contains approximately 3.6 million 3D human poses and corresponding image data from 11 professional actors (6 men and 5 women) performing in 17 different scenarios. The dataset includes pixel-level labels for 24 body parts of these actors and high-precision 3D joint data, which can be used to infer and calculate various human body dimensions.

### 4.2. Evaluation Metrics and Results

Among the methods introduced in Section 3, all methods used Mean Absolute Error (MAE) as one of the standards to evaluate model performance in the experiments. Most of the experiments recorded results for chest circumference, waist circumference, and hip circumference. Although these experiments were conducted on different datasets, they still allow for an objective evaluation of the models' performance to some extent. Table 3 shows the MAE results obtained from evaluating different models on different datasets.

**Table 3.** Results of the models tested on different datasets, using MAE as evaluation metrics.

| Model Name                         | Evaluation Metrics(MAE/mm) |       |       |              | Evaluation Dataset                        |
|------------------------------------|----------------------------|-------|-------|--------------|-------------------------------------------|
|                                    | Chest                      | Hips  | Waist | Overall      |                                           |
| Neural Anthropometer               | 25.22                      | 27.53 | 25.85 | 26.20        | SMPL Generated<br>HBW(Human Body<br>Wild) |
| Shapy                              | 65.00                      | 57.00 | 69.00 | 63.67        |                                           |
|                                    | 64.00                      | 74.00 | 98.00 | 78.67        |                                           |
| BMnet(Single View)                 | 33.95                      | 31.03 | 31.93 | 32.30        | MMTS                                      |
| BMnet(Multi<br>View+Height+Weight) | 15.92                      | 9.74  | 15.44 | <b>13.70</b> | BodyM                                     |
| GBWO-ENN                           | 13.80                      | 8.50  | 8.40  | <b>10.23</b> | ANSUR II                                  |
| IWOA-ENN-MC                        | 15.91                      | 17.16 | 17.85 | 16.97        | CAESAR                                    |
| UGA-BP-MC                          | 29.10                      | 23.00 | 20.00 | 24.03        | ANSUR II                                  |
| PC-BoDiEs                          | 32.90                      | -     | 22.90 | 27.90        | CAESAR                                    |
| Conv-BoDiEs                        | 25.70                      | -     | 16.50 | 33.95        | CAESAR                                    |

It can be learnt from the table that PC-BoDiEs and Conv-BoDiEs (using grayscale images) have an overall result calculated as the average of only two metrics due to the missing hip test values. It is worth mentioning that the original test results included a larger number of metrics; for comparison purposes, only a subset is used here. The original MAE recorded for Conv-BoDiEs (using grayscale images) was 46.40mm, and for PC-BoDiEs, it was 49.50mm. Similarly, due to the large number of measurement metrics in the original experimental results, the Neural Anthropometer model had an overall MAE of 20.89mm when all measurement metrics were considered.

#### 4.3. Evaluation Results Summary

As shown in Table 3, methods based on the ABSS algorithm for height, weight, and circumference values generally have higher accuracy. Among the image-based methods, both Neural Anthropometer and BMnet (Single View) exhibit good accuracy. It is worth noting that the latter shows a significant improvement in accuracy after incorporating multi-view images (not listed in the table), and its accuracy reaches an MAE of 13.70mm after including height and weight as metrics, almost comparable to the accuracy of GBWO-ENN, which does not rely on images. This highlights the important role of height and weight values and adversarial augmentation in improving the accuracy of human dimension estimation. It also provides a direction for future research: whether more accurate height and weight values can be obtained from images to assist in the measurement of more human dimension metrics, thereby enabling accurate measurement of multiple human dimensions using images alone.

## 5. Conclusion

As a research hotspot in the field of artificial intelligence in recent years, non-contact human body dimension estimation has significant practical implications in areas such as clothing design, health status estimation, and posture recognition. This paper categorizes human body dimension estimation methods based on the type of input data used: image-based human dimension estimation, estimation of height, weight, and circumference values using the ABSS algorithm, and estimation based on 3D human meshes or 3D point clouds. The paper summarizes the network models employed by these methods and their characteristics. It also introduces commonly used datasets and evaluation metrics in the field of human dimension estimation, analyzing the performance of different network models on various datasets.

Although significant research progress has been made in non-contact human body dimension estimation based on deep learning, several problems and challenges remain:

Quality of existing human image datasets needs improvement: The commonly used datasets contain some blurry images, and variations in gender and camera distance significantly impact the model's

estimation accuracy. Additionally, some datasets used in research have a relatively homogeneous sample type, such as the ANSUR II dataset, which uses body measurement data from U.S. Army soldiers. These factors can affect the accuracy of the model's human dimension estimation results.

Scarcity of extreme samples and lack of representativeness: Current research on human dimension estimation based on 3D human meshes or 3D point clouds mainly relies on SMPL synthetic data, which may lead to insufficient generalization ability in practical applications. Most training datasets also lack samples of extreme body types, potentially reducing the model's prediction accuracy for these rare body types. Using adversarial Body Simulators (ABS) may significantly increase the time and cost of generating samples.

Occlusion issues and user privacy concerns: When the subject is wearing loose-fitting clothing, changes in body contour may affect the model's prediction results. Collecting 3D human data or precise 2D human images often requires subjects to wear tight-fitting clothing or no clothing, which raises potential privacy issues. Additionally, for methods based on the ABSS algorithm, optimizing the adaptive segmentation method to obtain accurate values is an important factor to consider.

Computational power requirements for model deployment: As the performance of human dimension estimation network models improves, the computational power required for model deployment and operation in real-life and production scenarios may not be met, preventing the models from achieving their expected prediction results.

In summary, addressing the above issues requires the establishment of comprehensive, well-classified human dimensions and image datasets with clear indicators. Improving the quality of datasets and synthetic data, and creating representative datasets, are the main directions for future dataset optimization. Current research methods are constrained by deployment costs and computational power requirements, making them difficult to apply in personal practical use. Therefore, the development of lightweight human dimension estimation networks will be a primary research direction in the future. Building models that can adapt to various input data in different application scenarios and addressing potential issues in obtaining input data are key to the widespread application of human dimension estimation networks.

## References

- [1] Song, Y., Wirta, J., Kämäräinen, J. K. (2020). Anthropometric clothing measurements from 3d body scans. *Machine Vision and Applications*, 31(4): 7.
- [2] Dibra, E., Jain, H., Öztireli, C., Ziegler, R., Gross, M. (2016). Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks. In *International Conference on 3D Vision (3DV)*, 108–117.
- [3] Hu, X. R., Liu, J. W., Liu, J. P., et al. (2020). Multi-feature extraction and dimension measurement for dressed human bodies via adaptive body structure segmentation. *Computer Engineering*, 46(2): 238-246.
- [4] Hu, X. R., Liu, J. W., Liu, J. P., et al. (2021). Prediction of 3D human dimensions based on improved GA-BP-MC neural network. *Computer Engineering and Science*, 43(08): 1443-1453.
- [5] Ruiz, N., Bellver, M., Bolkart, T., Arora, A., Lin, M. C., Romero, J., Bala, R. (2022). Human body measurement estimation with adversarial augmentation. In *International Conference on 3D Vision (3DV)*, Prague, Czech Republic, September 12-16. IEEE: 219-230. DOI: 10.1109/3DV57758.2022.00031.
- [6] Yang, X. W., Li, Y. T., Han, X., et al. (2024). Human dimension prediction model integrating GBWO and ENN. *Computer Technology and Development*, 34(06): 132-139. DOI: 10.20165/j.cnki.ISSN1673-629X.2024.0064.
- [7] Hu, X. R., Liu, J. W. (2023). Prediction of human circumference dimensions based on improved WOA-ENN-MC model. *Computer Applications and Software*, 40(05): 190-199.

- [8] Bogo, A., Kanazawa, A., Lassner, P., et al. (2016). Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. *Proc of European Conference on Computer Vision (ECCV)*: 1.
- [9] Choutas, V., Müller, L., Huang, C. H. P., Tang, S., Tzionas, D., Black, M. J. (2022). Accurate 3D body shape regression using metric and semantic attributes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*: 2708-2718. DOI: 10.1109/CVPR52688.2022.00274.
- [10] Alldieck, T., Marcus, M., Xu, W. P., et al. (2018). Video-based reconstruction of 3D people models. *Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1.
- [11] Alldieck, T., Marcus, M. (2019). Learning to reconstruct people in clothing from a single RGB camera. *Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1.
- [12] Sun, Y. L., Miao, Y. W., Bao, C. (2019). Global registration cumulative error minimization based 3D human body reconstruction using RGB-D scanning data. *Journal of Computer-Aided Design & Computer Graphics*, 31(9): 1467-1476.
- [13] Kazhdan, M., Hoppe, H. (2013). Screened Poisson surface reconstruction. *ACM Transactions on Graphics*, 32(3): Article No. 29.
- [14] González, T. Y., Mayer, H. A. (2021). A neural anthropometer learning from body dimensions computed on human 3D meshes. *IEEE Symposium Series on Computational Intelligence (SSCI)*: 1-8.
- [15] Škorvánková, D., Riečický, A., Madaras, M. (2022). Automatic estimation of anthropometric human body measurements. *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*: 537-544. DOI: 10.5220/0010878100003124.