TOWARDS AUTOMATED GARMENT MEASUREMENTS IN THE WILD USING LANDMARK AND DEPTH ESTIMATION

By

Cristopher I. Zbavitel

December, 2024

Director of Thesis: David Hart, PhD

Major Department: Computer Science

ABSTRACT

This research introduces an innovative approach to automate garment measurements from photos to address the high return rates in the fashion industry due to inaccurate sizing. The presented method employs DepthAnything for depth estimation and Keypoint R-CNN for landmark estimation in a multi-stage approach, advancing previous methodologies by offering a scalable and accurate solution for the fashion industry. The technique is analyzed on the DeepFashion2 dataset and a custom set of images, and initial findings suggest promising avenues for reducing returns and enhancing the garment fitting processes.

TOWARDS AUTOMATED GARMENT MEASUREMENTS IN THE WILD USING LANDMARK AND DEPTH ESTIMATION

A Thesis

Presented to The Faculty of the Department of Computer Science

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Computer Science

By

Cristopher I. Zbavitel

December, 2024

Director of Thesis: David Hart, PhD

Thesis Committee Members:

Nic Herndon, PhD

Rui Wu, PhD

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In recent years, the fashion industry has experienced a seismic shift, primarily propelled by the surge in online shopping. While this transition offers consumers unparalleled variety and convenience at a global scale, it has also led to significant challenges, notably high return rates due to sizing inaccuracies. Inconsistent garment-sizing standards across different brands contribute to return rates as high as 25% in online apparel sales [16].

This research introduces a novel approach to automate garment measurements from photographs taken in uncontrolled environments ("in the wild"). Specifically, I developed a computer vision pipeline that combines advanced depth estimation and landmark detection to predict accurate garment measurements from a single image. It focuses on standard shirt measurements, aiming to reduce return rates and enhance the garment fitting process for both consumers and retailers.

To realize these objectives, a multi-stage approach was created. First, the 3D geometry information is extracted using advanced depth estimation models, notably the Depth Anything V2 [27, 28] neural network model. Second, the garment is segmented, and important landmarks are predicted using DeepLabV3 [6, 7] and Keypoint R-CNN [11, 24], respectively. This combined information provides a 3D object segmentation of the garment that can then be analyzed for measurements. An overview of this approach is given in Figure 1.1.

My contributions in this work are threefold. First, a multi-stage measurement pipeline was developed and it integrates state-of-the-art models—Depth Anything V2 for depth estimation and Keypoint R-CNN for landmark detection—to extract precise garment measure-

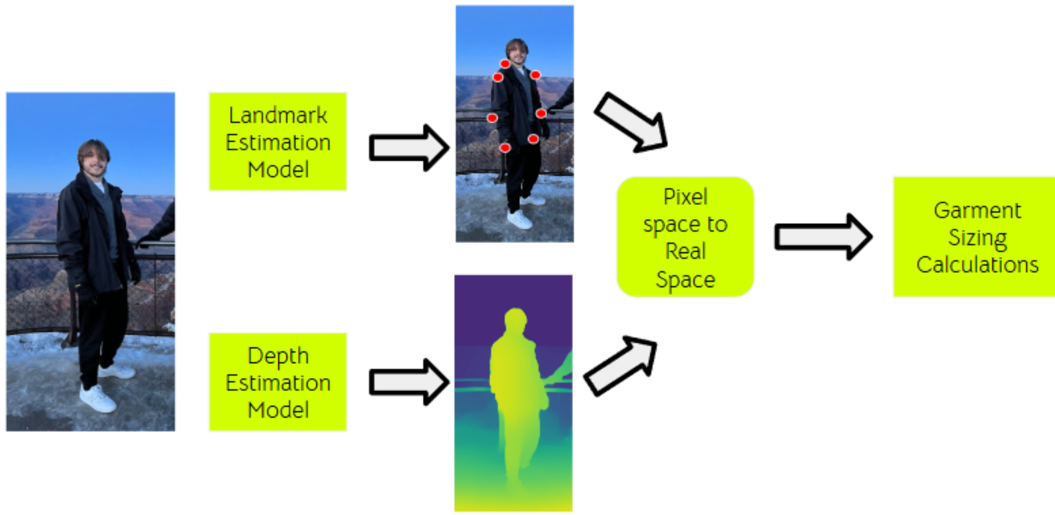

Figure 1.1: An overview of my approach. The pipeline integrates depth estimation, garment segmentation, and landmark detection to extract accurate garment measurements from a single image.

ments from single images. Second, an ablation study was conducted to assess the impact of depth estimation and keypoint detection on measurement accuracy. My findings reveal that using true depth with predicted keypoints yields the best performance, achieving a mean absolute error (MAE) of 5.08 cm in waist circumference estimation. Third, I validated my method across multiple datasets, including a custom dataset with ground truth measurements and the DeepFashion2 dataset, demonstrating scalability and robustness. On the custom dataset, my method achieves an MAE of 8.74 cm for sideseam measurements. These results indicate that the approach significantly improves upon existing techniques, providing a scalable and accurate solution for the fashion industry.

The remainder of this thesis is structured as follows: Chapter 2 offers a review of the literature on challenges in online fashion retail, existing solutions for garment sizing, and the role of technology in tackling these issues. Chapter 3 elaborates on the research methodology, detailing data collection methods, model development, and validation processes. Chapter 4 showcases the results and analysis of implementing and testing the system, discusses findings, and acknowledges limitations. Finally, Chapter 5 concludes the thesis, summarizing the research contributions and suggesting avenues for future exploration, as well as indicating the broader implications of these findings for both the industry and academic community.

Chapter 2

Literature Review

This literature review explores computational techniques pertinent to the automatic measurement of garment sizes using images. Both foundational work and state-of-the-art methodologies are presented, emphasizing their capabilities and limitations in addressing the problem of automated garment measurement from photographs.

1 Fashion Recommender Systems

Fashion recommender systems have employed sophisticated machine learning techniques to match garments with consumer profiles, leveraging accurate garment measurements. Modern systems utilize deep learning architectures to create personalized recommendations that account for individual fit and style preferences [18]. These systems enhance the shopping experience by analyzing user data to suggest items aligned with the consumer's tastes.

However, a significant limitation of fashion recommendation systems is the assumption that all important body measurements of the user are known. In practice, not all consumers are aware of their precise measurements at the time of purchase, which undermines the effectiveness of these systems. Additionally, they do not leverage visual information from images to infer measurements, restricting their ability to provide accurate size recommendations based on a user's appearance. Consequently, while fashion recommender systems excel in personalization, they are insufficient for the goal of predicting accurate garment sizes from a single image taken in uncontrolled environments.

2 Fashion and Computer Vision

The intersection of fashion and computer vision has led to advancements such as Virtual Try-On (VTO) systems [10, 29, 5], which allow users to visualize how garments will fit and look without physical interaction. VTO systems enhance the online shopping experience by providing a visual representation of garments on virtual avatars or directly on user images, leveraging computer vision techniques and advanced machine learning models.

Despite their contributions to user engagement, VTO systems primarily focus on visual appearance rather than precise sizing. They often lack accurate body measurements and do not perform quantitative analysis of garment dimensions, thereby failing to estimate the correct garment size for the user. Challenges also persist in refining image details, achieving realistic textures, handling occlusions, and ensuring accurate alignment. Therefore, while VTO systems represent a significant leap forward in online shopping, they do not address the need for automatic garment measurement from photographs, as they lack the capability to extract precise measurements necessary for size recommendations.

3 Segmentation

Accurate separation of the garment from the rest of the image is essential for measurement extraction, and segmentation algorithms are employed for this purpose. Modern segmentation models, such as U-Net [17], Detectron2 [25], Segmenter [21], and Segment Anything [13], have demonstrated high accuracy in isolating objects within images. In this work, I used DeepLabV3 [6] due to its effectiveness and simplicity in identifying clothing.

Segmentation models provide pixel-level masks that enable precise localization of the garment, which is crucial for subsequent analysis. However, segmentation alone does not provide information about the garment's three-dimensional shape or measurements. It lacks depth information and does not identify specific landmarks necessary for accurate measurement extraction. Therefore, while segmentation is a vital first step in the process, it cannot solve the problem of automatic garment measurement independently. Integration with depth estimation and landmark detection is necessary to obtain the 3D geometric information required for accurate sizing.

4 Human Pose Estimation and Keypoint Detection

After segmentation, detecting important landmarks on the garment that correlate to measurement areas, such as the waist or sideseam of a shirt, is essential. Keypoint detectors, often developed for human pose estimation, have evolved significantly, with models like HRNet [22] and ViTPose [26] demonstrating remarkable success in identifying anatomical landmarks. When fine-tuned on fashion datasets, these models show robustness in detecting fashion landmarks crucial for garment fitting [8, 20, 19].

In this work, I utilized Keypoint R-CNN [11], built on the Mask R-CNN pipeline, for its effectiveness in various applications, including human pose estimation and fashion landmark detection. Keypoint R-CNN is relatively straightforward to fine-tune on specific datasets like DeepFashion2 [8] and is supported by widely used frameworks such as PyTorch.

While keypoint detection models provide precise locations of key points necessary for measuring dimensions, they are typically designed for human pose estimation and may not accurately capture garment-specific landmarks, especially when the garment deviates from the body's contours. Their performance can be limited in complex poses or with occlusions, particularly in images taken in uncontrolled environments. Therefore, keypoint detection alone cannot provide accurate garment measurements without considering depth information. Integration with depth estimation is essential to reconstruct the three-dimensional geometry of the garment for accurate sizing.

5 Monocular Depth Estimation

Predicting depth information at each point is fundamental for generating the three-dimensional structure of the garment. Monocular depth estimators developed using deep learning approaches infer depth from a single image [9, 2, 15], enabling 3D understanding without specialized equipment. However, these approaches generally return relative depth values, accurate up to an unknown scaling and offset, which is insufficient for precise measurements.

The Depth Anything V2 model [27, 28] was employed, a state-of-the-art monocular depth estimator that provides absolute depth predictions in meters. This capability is crucial for this pipeline, as absolute and precise measurements are needed for each location on the garment. While Depth Anything V2 advances the field by offering absolute depth estimation, its accuracy may still be limited in complex scenes or with diverse garment textures. Performance may degrade in uncontrolled environments with varying lighting and background clutter.

Depth estimation models enable the reconstruction of the garment's three-dimensional geometry, but depth estimation alone cannot determine specific measurement points without landmark detection. Consequently, integrating depth estimation with keypoint detection is necessary to accurately extract garment measurements from images.

6 3D Mesh Reconstruction and Geodesic Measurement

Representing the garment as a three-dimensional geometry facilitates the calculation of measurements along its surface. Previous works have attempted to reconstruct meshes from discrete points, exploring approaches such as shape matching and object recognition based on contour [1]. The generated mesh allows for the calculation of geodesic distances, providing a natural measurement of garment sizes that adhere to the contours of the human body [3].

However, accurate geodesic distance calculation on 3D meshes presents several technical challenges. Irregular meshes often exhibit complex geometries with varying vertex densities, leading to potential inaccuracies. Topological variations such as holes and non-manifold edges complicate pathfinding algorithms [4]. Additionally, calculating geodesic distances on large or highly detailed meshes can be computationally expensive, requiring significant processing power and time [23]. Mesh reconstruction from single images is further challenged by occlusions, lack of texture, and limited viewpoints, leading to incomplete or inaccurate meshes.

Due to these challenges, particularly in uncontrolled environments with single-view images, mesh reconstruction may not be feasible or may result in unreliable measurements. In this work, I opted for using simple Euclidean distances between landmarks to avoid the irregularities and computational complexities associated with mesh reconstruction and geodesic distance calculations. While this approach may introduce some approximation errors, it offers a more practical and efficient solution for automatic garment measurement from images.

7 Automatic Garment Sizing

Several related works have attempted to measure garments from visual information using modern techniques. For instance, Li et al. [14] developed a system where garments were placed on mannequins at a consistent distance from a fixed camera, enabling precise measurement extraction. Similarly, Kim et al. [12] laid garments flat and captured images using an iPhone camera equipped with LiDAR to obtain depth information.

These methods achieve high accuracy in garment measurement by controlling the photographic environment, utilizing consistent distances, fixed camera positions, and additional sensors to reduce variability. However, they rely heavily on controlled conditions, which drastically simplifies the problem. Such setups are not practical for general consumer use or for images taken in the wild. The requirement for specialized equipment or specific environments limits their scalability and applicability in real-world scenarios.

Since these methods cannot operate effectively on arbitrary images captured in uncontrolled environments, they are insufficient for the goal of automatic garment measurement from single images taken by consumers. This work aims to overcome these limitations by developing a system capable of extracting accurate measurements without imposing strict constraints on the input photos, thereby advancing the field toward practical and scalable solutions.

Chapter 3

Methodology

In this research, I employed a combination of advanced computer vision techniques and machine learning models to estimate garment measurements from images captured in uncontrolled environments. This multi-stage pipeline integrates segmentation, keypoint detection, depth estimation, and measurement extraction to achieve accurate sizing predictions.

1 Pipeline Overview

An inherent difficulty arises in attempting to solve the automatic garment sizing task: there is no public dataset that connects images of people to their respective garment measurements. The lack of a complete dataset makes an end-to-end neural network approach implausible. Thus, I used a multi-stage mixed-methods pipeline that predicts keypoints and depth information to obtain garment measurements. Evaluating a multi-stage pipeline involves considering many models, datasets, and test cases to evaluate components independently.

An overview of this pipeline is shown in Figure 3.1 and is described in the following sections.

2 Models

2.1 Segmentation and Landmark Detection

The process begins by taking an input image and segmenting the clothing from the background. I used the DeepLabV3 [6] segmentation model to determine the pixels that represent

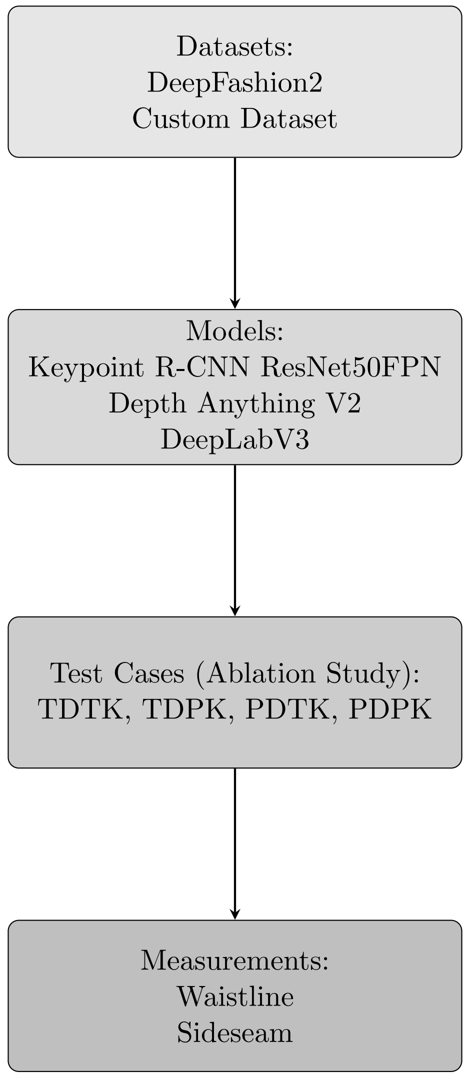

Figure 3.1: Overview of the methodology pipeline. This diagram illustrates the sequential stages of this approach, starting from input datasets, progressing through model components, evaluating different test cases in an ablation study, and concluding with the extraction of garment measurements. It highlights how each component interacts within the pipeline and the flow of data between stages.

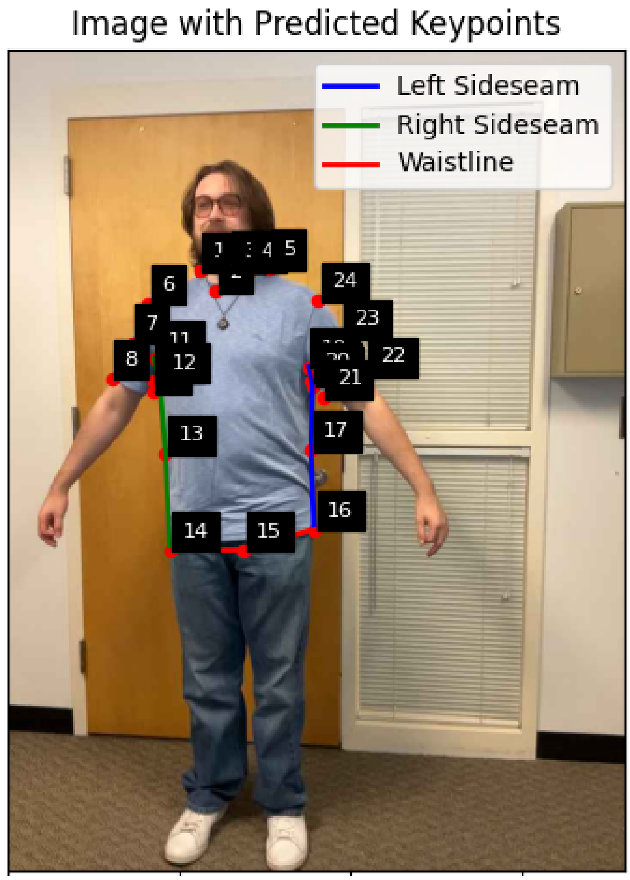

Figure 3.2: Example output of keypoint landmarks detected by Keypoint R-CNN on a garment image. The keypoints correspond to specific landmarks crucial for measurement extraction, such as the waistline and sideseams. This visualization demonstrates the model's ability to accurately identify important points on the garment, which are essential for calculating measurements.

the garment. Segmentation is crucial for isolating the garment and removing background noise that could interfere with subsequent processing.

Then, a keypoint detector identifies landmarks on the garments. Keypoint R-CNN [11], using a ResNet backbone, is employed for precise landmark estimation. This model is trained on the DeepFashion2 [8] images and focuses on detecting keypoints on garments. These keypoints correspond to specific anatomical landmarks necessary for measurement extraction, such as the waistline and sideseams. An illustration of these steps is provided in Figure 3.2.

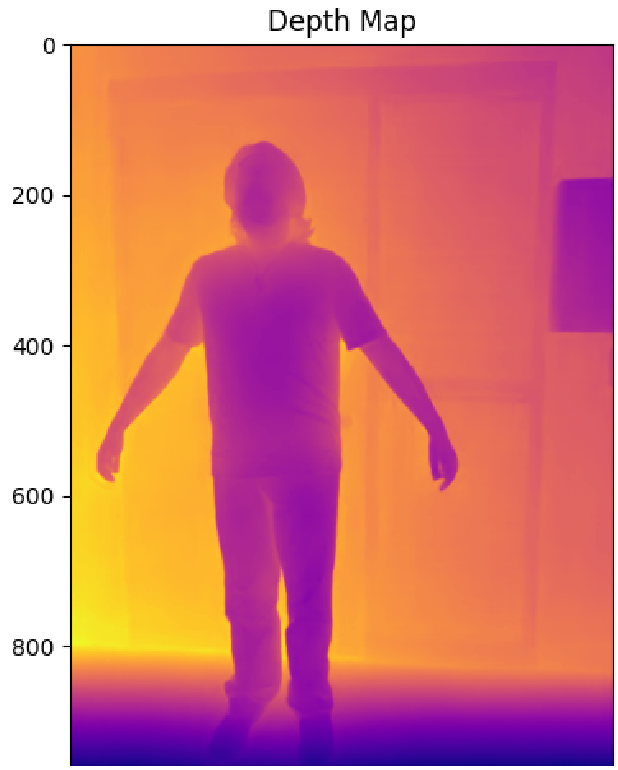

Figure 3.3: Example depth map output generated by Depth Anything V2 for a garment image. The depth map represents the distance of each pixel from the camera in meters, where lighter shades indicate closer regions and darker shades represent areas further away. Accurate depth estimation is critical for reconstructing the garment's 3D geometry and obtaining precise measurements.

2.2 Depth Estimation

The next step in the process is to obtain an accurate depth map of the image, describing how far away each point on the image is from the camera. This depth information is needed to determine the 3D geometry of the clothing to take accurate measurements. In some cases, true depth information is provided through a LiDAR camera, such as those on modern iPhone models. If depth information is not provided, the Depth Anything V2 model [28] is used to predict the depth for each pixel. Depth Anything V2 is selected for depth estimation due to its robust performance in generating dense depth maps from single images in absolute meters. It excels across diverse scenes, making it ideal for this application. An example depth map is shown in Figure 3.3.

2.3 Pixel Space to Real Space Conversion

The segmented image, keypoints, and depth information provide the essential data for reconstructing the garment in 3D space. However, as this information is initially captured in pixel space, it is necessary to convert the 2D keypoints into real-world 3D coordinates relative to the camera. This transformation is performed in two key steps.

First, to ensure stability in downstream calculations, any keypoints located outside the segmentation boundary are adjusted to the closest point within the segmented area. This adjustment addresses minor inaccuracies in segmentation and ensures that all keypoints correspond to valid garment pixels.

Second, with each keypoint's pixel coordinates (x', y') and the real-world depth value z from the depth map, the real-world coordinates (x, y, z) are calculated using the camera's intrinsic parameters:

$$x = \frac{(x' - w/2) \cdot z}{f_x} \tag{3.1}$$

$$y = \frac{(y' - h/2) \cdot z}{f_y} \tag{3.2}$$

where w and h are the image width and height, respectively, and f_x and f_y are the focal lengths in the x and y directions, derived from the camera's field of view (FOV). These calculations rescale the coordinates from pixel space to real-world space, compensating for perspective distortion and aligning each keypoint accurately within the 3D scene. By completing these transformations, the resulting (x, y, z) coordinates for each keypoint accurately reflect the garment's position in real-world space, which is essential for precise 3D garment reconstruction.

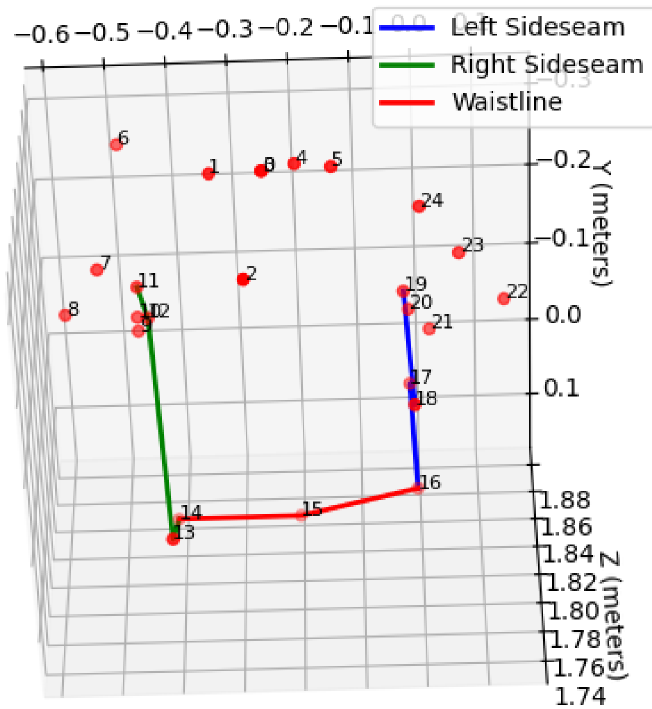

Figure 3.4: Visualization of the keypoints used to measure the final distances. The figure illustrates how Euclidean distances are calculated between specific keypoints to determine the waistline and sideseam measurements. By connecting these keypoints, I can compute the necessary measurements for garment sizing. Doubling the waistline measurement accounts for the full circumference, assuming body symmetry.

2.4 Measurement Extraction

Following the DeepFashion2 format, I evaluated images for waistline and sideseam measurements. Specific keypoints mark the waistline and sideseams. For simplicity and stability of results, the Euclidean distances are calculated between the necessary keypoints. For the waistline measurement, the value is doubled to account for the full circumference, assuming body symmetry. A visualization of this final step is given in Figure 3.4.

Due to challenges in reconstructing a full 3D mesh with only front-view data, I used Euclidean distance calculations as an approximation for geodesic measurements in this pipeline.

3 Testing and Evaluation

As described in Section 2, garment measurements can be extracted for a given input image. To evaluate the impact of each component in the pipeline, I conducted an ablation study using four different test cases:

1. **True Depth with True Keypoints (TDTK)**: Uses ground truth depth data and manually annotated keypoints.

2. **True Depth with Predicted Keypoints (TDPK)**: Uses ground truth depth data and keypoints predicted by Keypoint R-CNN.

3. **Predicted Depth with True Keypoints (PDTK)**: Uses depth predicted by Depth Anything V2 and manually annotated keypoints.

4. **Predicted Depth with Predicted Keypoints (PDPK)**: Uses both depth and keypoints predicted by the models, representing the full pipeline without any ground truth data.

This ablation study allows us to assess the impact of depth estimation and keypoint detection on measurement accuracy individually and collectively. By comparing the results across these configurations, I identified which components contribute most to the accuracy and consistency of the measurements.

For each test case, I calculated the Mean Absolute Error (MAE) and Standard Deviation (SD) of the measurements compared to the ground truth measurements from my custom dataset. In cases without ground truth measurements (e.g., DeepFashion2 samples), the consistency of the model's predictions were analyzed across subjects by computing the mean and standard deviation of the results.

4 Datasets

As stated previously, no public dataset that goes from image to garment measurements exists. Therefore, I utilized multiple datasets to train and evaluate different components of the pipeline, as well as to test the overall system under various conditions.

4.1 DeepFashion2

DeepFashion2 [8] is utilized primarily for training the keypoint detection models due to its comprehensive collection of fashion images annotated for segmentation and landmark detection. Its extensive use in fashion-related computer vision tasks and its COCO-like structure make it highly suitable for this project's purposes.

15-Image Focused Sample

To evaluate the method under ideal conditions, a focused sample of 15 images was sampled from the DeepFashion2 dataset. These images were chosen to have consistent camera angles, full visibility of garments, and minimal occlusions, resembling the conditions of the custom dataset. This sample allows us to assess the performance of the method in favorable scenarios.

573-Image Random Sample

To test the robustness and generalizability of the method, I also evaluated it on a random sample of 573 images from DeepFashion2 without any filtering. This sample includes images with diverse poses, camera angles, occlusions, and varying image qualities, representing real-world conditions. This evaluation helps identify potential limitations of the method in uncontrolled environments.

4.2 Custom Dataset with iPad and RealSense

We created a small custom dataset of images using an iPad with LiDAR and a RealSense depth camera. The dataset consists of pictures of two subjects, with six samples taken per subject, totaling 12 images. The fashion keypoints were hand-annotated on each photo, and true garment measurements were taken of each participant using a tailor's tape. Outputs from the custom dataset are shown in Figure 3.5.

This custom dataset serves multiple purposes:

- **Baseline Evaluation**: Provides ground truth measurements for evaluating the absolute accuracy of my method.
- **Controlled Conditions**: Allows us to test the pipeline in a controlled environment with high-quality depth data and precise keypoint annotations.
- **Ablation Study**: Enables us to perform an ablation study by testing different configurations (e.g., using true depth vs. predicted depth).

By utilizing these datasets, I ensure a comprehensive evaluation of the method across various conditions and configurations.

5 Tools and Implementation

This implementation utilizes Python's powerful libraries, including OpenCV, NumPy, and PyTorch. These tools support various stages of image processing, depth estimation, and keypoint estimation, ensuring a robust and efficient workflow. The models were trained and executed on hardware equipped with NVIDIA GPUs to accelerate computations.

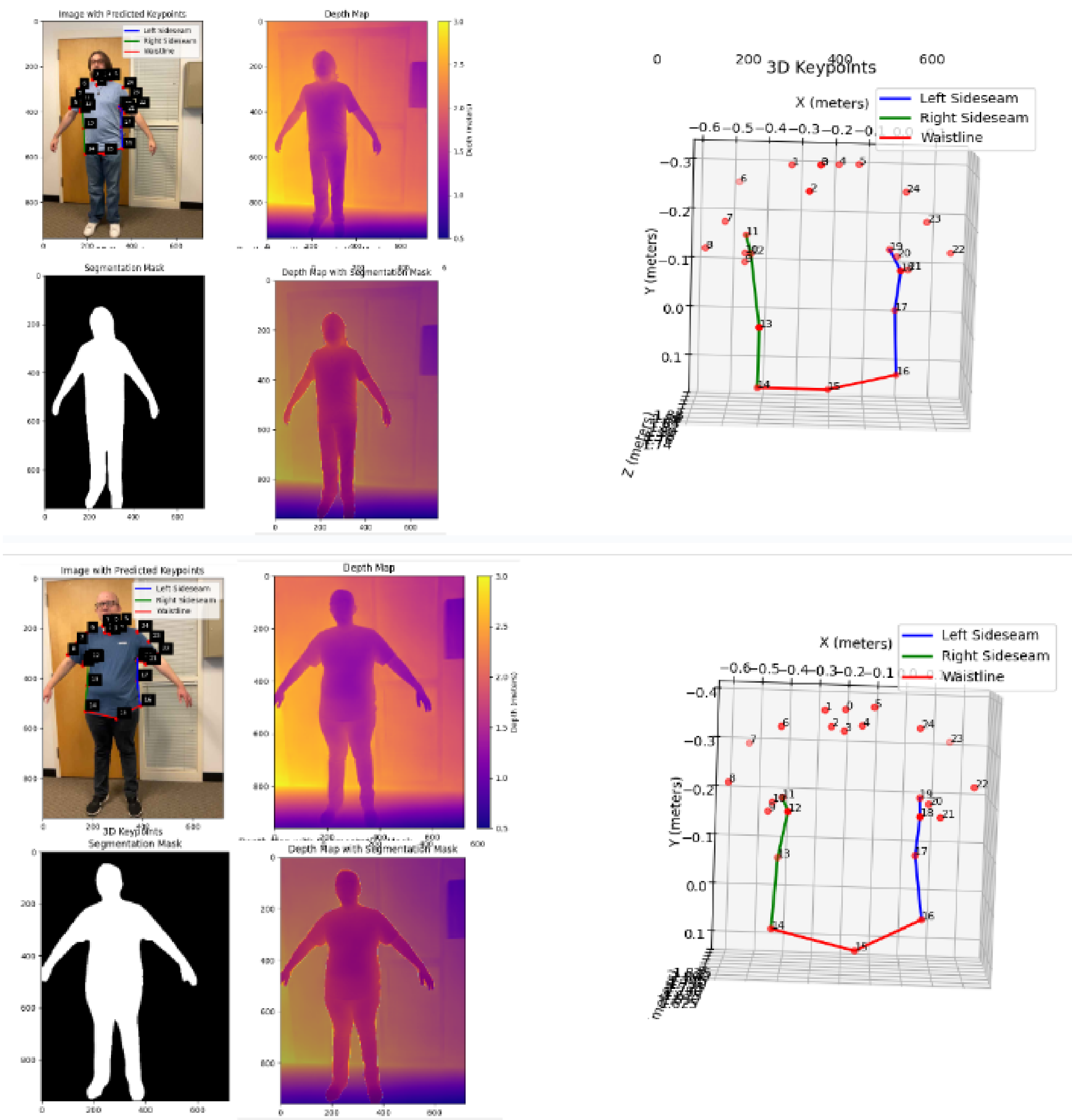

Figure 3.5: Outputs from the custom dataset for Person A (top) and Person B (bottom). Each image displays the detected key points and segmentation masks overlaid on the subjects. Depth was measured precisely using a LiDAR-equipped iPad camera and key points were hand-labeled for this dataset.

Chapter 4

Results

This chapter presents the results of my approach on the datasets outlined in Chapter 3. I conducted an ablation study to evaluate the impact of depth estimation and keypoint detection on measurement accuracy, using different configurations of the pipeline. The results for each dataset are discussed below.

1 Custom Dataset

1.1 Ground Truth Measurements

For the custom dataset, precise garment measurements were collected using a tailor's tape to serve as ground truth for evaluating the accuracy of these methods. The measurements for each subject are provided in Table 4.1.

Table 4.1: Ground truth measurements of subjects in the custom dataset obtained using a tailor's tape. These measurements serve as the baseline for evaluating the accuracy of my measurement extraction methods.

Subject	**Full Waist (m)**	**Half Waist (m)**	**Sideseam (m)**
Person A	0.99	0.495	0.43
Person B	1.23	0.615	0.48

1.2 Ablation Study on the Custom Dataset

Our approach depends on two vital components: a depth estimator and a keypoint estimator. To assess the impact of each component on measurement accuracy, I conducted an ablation study using four configurations of the method:

- **TDTK** (True Depth, True Keypoints): Both depth and keypoints are ground truth.
- **TDPK** (True Depth, Predicted Keypoints): Ground truth depth with predicted keypoints.
- **PDTK** (Predicted Depth, True Keypoints): Predicted depth with ground truth keypoints.
- **PDPK** (Predicted Depth, Predicted Keypoints): Both depth and keypoints are predicted.

By comparing these configurations, I can identify the contributions of depth estimation and keypoint detection to the overall measurement accuracy.

The average measurements over six samples for both participants are provided under all four configurations. The results are presented in Tables 4.2 and 4.3.

Table 4.2: Person A's estimated measurements under different configurations. This table shows the average sideseam and waist circumference measurements, allowing comparison across configurations to assess the impact of depth and keypoint prediction.

	Average Sideseam (m)	**Average Waist Circumference (m)**
TDTK	0.2877	0.8185
TDPK	0.3540	0.9637
PDTK	0.3015	0.7454
PDPK	0.3311	0.8354
Ground Truth	0.43	0.99

Table 4.3: Person B's estimated measurements under different configurations. This table allows for analysis of how depth and keypoint prediction affect measurement accuracy for a different subject.

	Average Sideseam (m)	**Average Waist Circumference (m)**
TDTK	0.2965	1.0006
TDPK	0.3258	1.0809
PDTK	0.2891	1.0450
PDPK	0.3332	1.0505
Ground Truth	0.48	1.23

To quantify the accuracy of the measurements, I calculated the Mean Absolute Error (MAE) and Standard Deviation (SD) for each configuration. The error metrics for both subjects are summarized in Tables 4.4 and 4.5.

Table 4.4: Person A's error metrics for sideseam and waist circumference measurements across the six samples. This table presents the MAE and SD, highlighting the measurement consistency and accuracy under each configuration.

	Sideseam MAE (m)	Sideseam SD (m)	Waist MAE (m)	Waist SD (m)
TDTK	0.1423	0.0374	0.1715	0.0447
TDPK	0.0760	0.0243	0.0263	0.0217
PDTK	0.1285	0.0361	0.2446	0.0361
PDPK	0.0989	0.0224	0.1546	0.0387

Table 4.5: Person B's error metrics for sideseam and waist circumference measurements across the six samples. The MAE and SD values illustrate the effects of depth and keypoint predictions on measurement accuracy for this subject.

	Sideseam MAE (m)	Sideseam SD (m)	Waist MAE (m)	Waist SD (m)
TDTK	0.1835	0.0331	0.2294	0.0551
TDPK	0.1542	0.0224	0.1490	0.0784
PDTK	0.1909	0.0639	0.1850	0.0584
PDPK	0.1468	0.0316	0.1795	0.0714

1.3 Discussion of Custom Dataset Results

Overview of Measurement Performance

Our analysis of measurements from the two subjects across the four configurations reveals significant patterns in the accuracy and reliability of the measurement approach. The most notable finding is that the **TDPK** configuration consistently produced the best results for both subjects. For Person A, the waist circumference MAE was 0.0263 m (approximately 1.04 inches), and the sideseam MAE was 0.0760 m (approximately 2.99 inches). For Person B, the waist circumference MAE was 0.1490 m (approximately 5.87 inches), and the sideseam MAE was 0.1542 m (approximately 6.07 inches).

Comparative Analysis of Configurations

Configuration 1: TDTK (True Depth, True Keypoints)

Contrary to my initial expectations, using ground truth for both depth and keypoints did not yield the best results. For Person A, this configuration produced a waist circumference MAE of 0.1715 m and a sideseam MAE of 0.1423 m. Person B's measurements showed even higher error rates, with waist circumference MAE of 0.2294 m and sideseam MAE of 0.1835 m. This unexpected outcome suggests that manual keypoint annotations may introduce inconsistencies affecting measurement accuracy.

Configuration 2: TDPK (True Depth, Predicted Keypoints)

This configuration emerged as the most accurate, with both subjects showing their lowest MAE values. The superior performance of predicted keypoints when combined with true depth data indicates that the keypoint prediction algorithm provides more consistent reference points than manual annotation.

Configuration 3: PDTK (Predicted Depth, True Keypoints)

The introduction of predicted depth data led to a significant increase in error rates. For Person A, the waist circumference MAE increased to 0.2446 m, and the sideseam MAE was 0.1285 m. For Person B, the waist circumference MAE was 0.1850 m, and the sideseam MAE was 0.1909 m. This degradation in accuracy highlights the importance of accurate depth information in the measurement pipeline.

Configuration 4: PDPK (Predicted Depth, Predicted Keypoints)

When using predicted values for both parameters, the results showed moderate improvement over Configuration 3. For Person A, the waist circumference MAE was 0.1546 m, and the sideseam MAE was 0.0989 m. For Person B, the waist circumference MAE was 0.1795 m, and the sideseam MAE was 0.1468 m. This suggests some robustness in the prediction algorithms, as the combination of predicted values did not result in compounded errors.

Subject-Specific Variations

Person A's Measurements

Person A's measurements generally showed lower error rates, particularly in the TDPK configuration. The standard deviations across measurements were also relatively consistent, indicating good measurement stability. However, the consistent underestimation across all configurations suggests a systematic bias in this measurement approach.

Person B's Measurements

Person B's measurements exhibited higher MAE values across all configurations, particularly in waist circumference measurements. The standard deviations were slightly higher, suggesting more variability in measurement accuracy. This difference might be attributed to variations in body shape affecting keypoint detection accuracy, different clothing choices impacting depth measurements, or potential differences in pose consistency during measurement.

2 DeepFashion2 Dataset Results

To assess the generalizability and consistency of my method in varied image conditions, two samples from the DeepFashion2 dataset were evaluated:

- **15-Image Focused Sample**: A curated set of images with consistent camera angles and full garment visibility, resembling ideal conditions.
- **573-Image Random Sample**: An unfiltered, randomly selected set of images representing real-world conditions with diverse poses, occlusions, and image qualities.

As ground truth depth data is not available for this dataset, the goal here was to ensure realistic and consistent measurement estimates across different images and angles.

2.1 15-Image Focused Sample Results

We tested the method using both the PDPK and PDTK configurations on the focused sample. The results are presented in Tables 4.6 and 4.7.

Table 4.6: 15-Image Sample Results - Predicted Depth with Predicted Keypoints (PDPK). This table shows the sideseam and waist circumference measurements for each image, demonstrating the model's performance under ideal conditions with fully predicted data. The mean and standard deviation indicate the consistency of the measurements across the sample.

Image ID	Sideseam (in)	Waist Circumference (in)
000373	11.70	25.44
000609	12.52	21.74
000610	18.59	32.34
000611	9.44	24.80
000612	7.64	23.54
000638	15.44	42.22
000639	10.96	30.46
000643	12.15	31.28
000647	16.64	43.88
000889	10.88	31.88
000891	11.15	25.10
000894	8.10	19.26
000898	14.90	44.96
000911	16.31	36.28
016915	17.47	30.10
Mean	12.87	31.20
SD	3.13	7.92

Analysis of Focused Sample Results

For the PDPK configuration, the mean sideseam length is approximately 12.87 inches with a standard deviation of 3.13 inches, while the mean waist circumference is about 31.20 inches with a standard deviation of 7.92 inches. These relatively low standard deviations suggest that the measurements are fairly consistent across the sampled images when using predicted keypoints.

In contrast, the PDTK configuration yields a higher mean sideseam length of 18.01 inches and a mean waist circumference of 38.54 inches, with larger standard deviations of 6.99 inches and 17.45 inches, respectively. The greater variability in the PDTK measurements indicates less consistency when using true keypoints with predicted depth.

These results suggest that the PDPK method produces more consistent and potentially more reliable measurements compared to the PDTK method. One possible explanation is

Table 4.7: 15-Image Sample Results - Predicted Depth with True Keypoints (PDTK). This table presents measurements using true keypoints, allowing comparison with the PDPK method to assess the impact of keypoint prediction. The larger standard deviations indicate greater variability in the measurements.

Image ID	**Sideseam (in)**	**Waist Circumference (in)**
000373	8.83	23.04
000609	14.97	23.92
000610	14.61	34.56
000611	9.38	24.46
000612	7.89	21.12
000638	17.52	48.72
000639	15.06	28.00
000643	16.03	47.38
000647	31.79	88.10
000889	24.08	33.54
000891	18.95	28.84
000894	14.55	38.06
000898	29.88	42.04
000911	20.69	48.50
016915	25.89	35.38
Mean	18.01	38.54
SD	6.99	17.45

that the predicted keypoints are inherently aligned with the model's depth predictions, leading to measurements that are internally consistent within the model's framework. Conversely, using true keypoints may introduce discrepancies between the actual keypoint locations and the model's predicted depth, resulting in greater variability.

2.2 573-Image Random Sample Results

We evaluated the model's performance on an unfiltered random sample of 573 images to assess its robustness in real-world conditions. The results are summarized below:

Predicted Depth with Predicted Keypoints (PDPK):

- **Mean Sideseam**: 50.11 inches
- **Sideseam SD**: 24.90 inches
- **Mean Waist Circumference**: 74.84 inches

- **Waist Circumference SD**: 41.56 inches

Predicted Depth with True Keypoints (PDTK):

- **Mean Sideseam**: 32.20 inches
- **Sideseam SD**: 20.64 inches
- **Mean Waist Circumference**: 65.74 inches
- **Waist Circumference SD**: 40.54 inches

Analysis of Random Sample Results

The high standard deviations and unusually large mean measurements indicate significant variability and potential issues with the model's predictions in uncontrolled environments. Factors contributing to these results include pose variations, occlusions, garment diversity, and depth prediction errors. The PDTK method showed slightly lower mean measurements and reduced variability compared to the PDPK method.

3 Discussion of Results

3.1 Potential Sources of Error

A thorough analysis of the sources of errors in this measurement pipeline reveals several contributing factors:

Image Quality Factors

Variability in image quality, inconsistencies in hand-annotated keypoints, variability in camera angles, lighting conditions, subject pose, and image artifacts can negatively affect the accuracy of keypoint and depth estimation, leading to measurement errors.

Model Limitations

The limitations inherent in the models used for depth estimation and keypoint detection constitute another major source of error. Predicted depth data often exhibit higher errors compared to true depth data. Keypoint detection precision is also critical, and misidentified or misplaced keypoints can result in incorrect measurements.

System Parameters

Variability in camera parameters, such as focal length and field of view, can lead to inconsistent depth estimations and affect the accuracy of real-world measurements. Calibration errors in the camera system can introduce systematic biases across all measurements.

3.2 Analysis of Errors

The consistent underestimation observed across test cases suggests the presence of systematic biases in the model or methodology. These biases are likely due to oversimplified measurement assumptions, such as assuming body symmetry. Random errors can arise from variability in subject pose, clothing, and environmental conditions. Outlier cases indicate that the model may struggle with generalization beyond typical training scenarios.

3.3 Limitations in Current Implementation

Key limitations in the current implementation include:

- The assumption of body symmetry in waist circumference measurements, potentially contributing to systematic errors.
- The use of Euclidean distances rather than geodesic distances, leading to underestimation on curved surfaces.
- Inconsistencies between true and predicted keypoint locations affecting measurement accuracy.
- Dependency on accurate depth data and camera calibration.

3.4 Impact of Measurement Methodology

My methodology shows promise as a baseline approach, especially with predicted keypoints and true depth data. However, to improve accuracy and reliability, future work should focus on:

- Incorporating geodesic distance calculations to account for body contours.
- Enhancing depth estimation algorithms for better precision.
- Refining keypoint detection methods.
- Addressing calibration issues and improving handling of body asymmetry.

By addressing the identified limitations and incorporating advanced measurement techniques, the accuracy and reliability of the automated body measurement system can be significantly improved.

Chapter 5

Conclusion

In this thesis, a technique has been presented for predicting garment sizes from a single photo of a person, achieving a Mean Absolute Error (MAE) as low as 0.0263 meters (approximately 1.04 inches) for waist circumference measurements under controlled conditions. The multi-stage method works even in the absence of ground truth measurement data, establishing an effective baseline approach while highlighting key areas for improvement in future methods and datasets. Additionally, I have provided a thorough analysis of the consistency of depth and keypoint estimators and their effects on this approach through an ablation study.

1 Custom Dataset Conclusions

The results of my study demonstrate the potential of automated measurement methods while also highlighting current limitations. The **True Depth with Predicted Keypoints (TDPK)** approach emerges as the most promising, with waist circumference errors as low as 0.0263 meters (approximately 5.3% error) for Person A and 0.1490 meters (approximately 12.1% error) for Person B. Sideseam measurements also showed improved accuracy under this configuration, with MAE values of 0.0760 meters (approximately 2.99 inches) for Person A and 0.1542 meters (approximately 6.07 inches) for Person B. However, the consistent underestimation across all methods suggests the presence of systematic biases that need to be addressed. Future work could increase measurement accuracy through improved keypoint density, better depth estimation, and more sophisticated body modeling approaches.

The Depth Anything V2 model may not generalize well across varying garment textures or lighting conditions, especially when predicting fine-grained depth details for thin fabrics or occluded areas. This limitation underscores the need for more robust depth estimation models capable of handling diverse visual scenarios.

While Keypoint R-CNN is robust for general-purpose human pose estimation, garment-specific keypoint estimation may require a more tailored approach. The model might struggle with older photos, intricate poses and camera angles, or when keypoints are occluded (e.g., when hands cross over the waistline). These challenges indicate that specialized keypoint detection models could improve measurement accuracy for clothing applications.

The **Custom Dataset**, with controlled conditions and ground truth measurements, demonstrated that this approach is viable under ideal circumstances. The low errors in both waist circumference and sideseam measurements, especially when using true depth and predicted keypoints, validate the effectiveness of my method in a controlled environment.

2 DeepFashion2 Conclusions

For the **DeepFashion2** dataset, the 15-image focused sample yielded much more consistent results compared to the 573-image random sample. Using the **Predicted Depth with Predicted Keypoints (PDPK)** method on the focused sample, we obtained a mean sideseam length of 12.87 inches with a standard deviation of 3.13 inches, and a mean waist circumference of 31.20 inches with a standard deviation of 7.92 inches. In contrast, the random sample showed significant variability, with a mean sideseam length of 50.11 inches and a standard deviation of 24.90 inches, and a mean waist circumference of 74.84 inches with a standard deviation of 41.56 inches. The high variability in the random sample reflects the challenge of predicting accurate keypoints and depth for images with diverse poses, occlusions, and varying image qualities. This finding highlights the importance of dataset quality and diversity in developing reliable measurement systems.

3 Future Research Directions

3.1 Technical Improvements

To address the limitations identified in my study, several technical improvements are proposed. Enhancing depth estimation is a critical area for future research. Integrating advanced depth estimation models or using multi-view stereo techniques could improve depth accuracy by combining information from multiple perspectives. Implementing deep learning-based depth refinement methods may also enhance the precision of depth maps. Additionally, exploring sensor fusion approaches—such as combining data from RGB cameras and depth sensors—could lead to more accurate and robust depth estimations.

Advanced body modeling is another essential area for improvement. Developing asymmetric body models can account for individual variations in body shape, leading to more accurate measurements. Integrating statistical body shape priors can provide a foundation for modeling typical body shapes and variations. Implementing non-rigid deformation models allows for the representation of body movements and deformations, which is important for capturing accurate measurements in dynamic scenarios.

Integrating geodesic distance calculations into the measurement process is also a promising direction. Investigating real-time surface reconstruction methods can enable the calculation of distances along the body's surface rather than straight lines, providing more accurate measurements. Developing efficient geodesic distance approximations can make these calculations computationally feasible for real-time applications.

3.2 Experimental Extensions and Improved Datasets

Further experimental work is needed to validate and extend these findings. Expanding the dataset to include a more diverse range of body types, poses, and clothing styles is essential for improving the generalizability of the measurement system. Collecting data with precise camera calibration can also mitigate systematic biases observed in the measurements.

Integrating multiple sensor types, such as depth cameras, infrared sensors, and LiDAR, can enhance data richness and improve measurement accuracy.

Refining the measurement process can also enhance accuracy. Developing probabilistic measurement models can account for uncertainties in the data and provide confidence intervals for the measurements. Investigating multi-image measurement strategies can improve accuracy by considering multiple predictions at different points or from different perspectives.

3.3 System Integration

For practical applications, integrating the measurement system into real-world environments is crucial. Developing user-friendly interfaces will make the system accessible to non-technical users, such as retail employees or consumers. Implementing privacy-preserving protocols is essential to ensure that users' body measurements are handled securely and responsibly, addressing concerns about data privacy and compliance with regulations. Integrating the system with existing retail systems will facilitate its adoption in e-commerce platforms or in-store solutions, enhancing the shopping experience.

4 Industry Implications and Applications

4.1 Fashion and Retail Applications

Our findings demonstrate both the potential and current limitations of automated body measurement systems in the fashion industry. With the TDPK approach achieving MAE values as low as 2.63 cm for waist measurements, this system approaches but does not yet meet the typical industry standard of ± 2.5 cm tolerance for garment fitting. Despite this, several promising applications emerge.

In the context of virtual try-on systems, this measurement system can serve as a pre-screening tool for approximate size recommendations. By integrating with existing e-commerce platforms, it can enhance the online shopping experience by providing customers with more accurate sizing information. This has the potential to reduce return rates in online clothing sales by minimizing the mismatch between customer expectations and the actual fit of garments.

The system can also be applied in custom tailoring. It can provide initial measurements for made-to-measure clothing, streamlining the process for both customers and tailors. Additionally, it can be used for automated progress tracking in fitness clothing, allowing users to monitor changes in their body measurements over time.

Moreover, this measurement system can contribute to inventory optimization in the fashion industry. By collecting data on customer body measurements, retailers can perform size distribution analysis to better understand the proportions of different sizes needed. This information can inform regional fit preference mapping, allowing retailers to adjust their stock based on the preferences of specific markets.

4.2 Healthcare Applications

While these current accuracy levels may not meet medical standards, the system shows promise for healthcare applications. It can be utilized for non-invasive patient monitoring, providing a convenient way to track changes in body measurements without the need for direct contact or clinical settings. Long-term body shape change tracking can be valuable for monitoring the progress of patients undergoing treatments that affect body composition, such as weight loss programs or rehabilitation.

4.3 Technical Feasibility and Cost Analysis

From a technical feasibility and cost perspective, the system's reliance on standard RGB-D cameras makes it commercially viable. The hardware costs are relatively low, with installations potentially costing under $500. This affordability makes the system accessible to a wide range of users, including small businesses and individual consumers. The system also requires minimal computational resources for real-time processing, making it suitable for deployment on standard computing hardware without the need for specialized equipment.

BIBLIOGRAPHY

[1] Belongie, S., Malik, J., and Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 4 (2000), 509–522.

[2] Bhat, S. F., Alhashim, I., and Wonka, P. Adabins: Depth estimation using adaptive bins. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 4009–4018.

[3] Calla, L. A. R., Perez, L. J. F., and Montenegro, A. A. A minimalistic approach for fast computation of geodesic distances on triangular meshes. *Computers And Graphics XX*, XX (2023), XX–XX.

[4] Campen, M., and Kobbelt, L. Exact and approximate redistancing for level set methods. *Computer Graphics Forum 32*, 8 (2013), 3–12.

[5] Chen, C.-Y., Chen, Y.-C., Shuai, H.-H., and Cheng, W.-H. Size does matter: Size-aware virtual try-on via clothing-oriented transformation try-on network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 7513–7522.

[6] Chen, L.-C. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017).

[7] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence 40*, 4 (2017), 834–848.

[8] Ge, Y., Zhang, R., Wu, L., Wang, X., Tang, X., and Luo, P. A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. *CVPR* (2019).

[9] Godard, C., Mac Aodha, O., and Brostow, G. J. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 270–279.

[10] Han, X., Wu, Z., Wu, Z., Yu, R., and Davis, L. S. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 7543–7552.

[11] He, K., Gkioxari, G., Dollár, P., and Girshick, R. Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2980–2988.

[12] Kim, S., Moon, H., Oh, J., Lee, Y., Kwon, H., and Kim, S. Automatic measurements of garment sizes using computer vision deep learning models and point cloud data. *Applied Sciences 12*, 5286 (2022).

[13] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. Segment anything. *arXiv:2304.02643* (2023).

[14] Li, C., Xu, Y., Xiao, Y., Liu, H., Feng, M., and Zhang, D. Automatic measurement of garment sizes using image recognition. In *Proceedings of the 1st International Conference on Graphics and Signal Processing* (2017), pp. 30–34.

[15] Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., and Koltun, V. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence XX*, XX (2020), 1–14.

[16] Research, C. The true cost of apparel returns: Alarming return rates require loss minimization solutions, 2023. Accessed: 2024-07-17, Published: 2023-04-04.

[17] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (2015), Springer, pp. 234–241.

[18] Shirkhani, S., Mokayed, H., Saini, R., et al. Study of ai-driven fashion recommender systems. *SN Computer Science 4*, 514 (2023).

[19] Sidnev, A., Krapivin, A., Trushkov, A., Krasikova, E., Kazakov, M., and Viryasov, M. Deepmark++: Real-time clothing detection at the edge. In *proceedings of the IEEE/CVF winter conference on applications of computer vision* (2021), pp. 2980–2988.

[20] Sidnev, A., Trushkov, A., Kazakov, M., Korolev, I., and Sorokin, V. Deepmark: One-shot clothing detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (2019), pp. 0–0.

[21] Strudel, R., Garcia, R., Laptev, I., and Schmid, C. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (2021), pp. 7262–7272.

[22] Sun, K., Xiao, B., Liu, D., and Wang, J. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 5693–5703.

[23] Surazhsky, V., Surazhsky, T., Kirsanov, D., Gortler, S. J., and Hoppe, H. Fast exact and approximate geodesics on meshes. *ACM transactions on graphics (TOG) 24*, 3 (2005), 553–560.

[24] torchvision, P. *Keypoint R-CNN Visualization Utilities*. Pytorch, 2023. Accessed: 2024-07-17.

[25] Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

[26] Xu, Y., Zhang, J., Zhang, Q., and Tao, D. ViTPose: Simple vision transformer baselines for human pose estimation. In *Advances in Neural Information Processing Systems* (2022).

[27] Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., and Zhao, H. Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR* (2024).

[28] Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., and Zhao, H. Depth anything v2. *arXiv:2406.09414* (2024).

[29] Yu, R., Wang, X., and Xie, X. Vtnfp: An image-based virtual try-on network with body and clothing feature preservation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2019).

Appendix A: Full Results

1 Custom Ground Truth Measurements

Tailor's tape measurements were used as ground truth for evaluating the accuracy of our methods:

Table 1: Ground Truth Measurements (Tailor's Tape)

Subject	Full Waist (m)	Half Waist (m)	Sideseam (m)
Person A	0.99	0.495	0.43
Person B	1.23	0.615	0.48

2 Person A's Results: Side-seam and Waistline Measurements

2.1 Test Case 1: True Depth - True Key points

Table 2: Person A's Measurements - Test Case 1

Sample	Sideseam Distance (m)	Sideseam Distance (in)	Waistline Distance (m)	Waistline Distance (in)
1	0.2694	10.61	0.4011	15.79
2	0.2844	11.20	0.4266	16.80
3	0.2670	10.51	0.3979	15.66
4	0.2682	10.56	0.3805	14.98
5	0.2828	11.13	0.4438	17.47
6	0.3542	13.94	0.4057	15.97

Statistical Analysis for Test Case 1

- **Assumptions and Methodology:**
  - **Waistline Measurement Comparison:** The ground truth waist measurement is the full waist circumference (0.99 m). Since the model provides a linear waistline distance (front width), we estimate the waist circumference by doubling the waistline distance (assuming body symmetry). Estimated Waist Circumference $= 2 \times$ Waistline Distance.
  - **Sideseam Measurement Comparison:** We compare the model's sideseam distances directly with the ground truth sideseam length (0.43 m).

- **Calculations:**

Estimated Waist Circumferences and Errors:

Table 3: Person A's Waist Circumference Errors - Test Case 1

Sample	**Estimated Waist Circumference (m)**	**Error (m)**	**Absolute Error (m)**	**Squared Error (m^2)**
1	0.8022	-0.1878	0.1878	0.0353
2	0.8532	-0.1368	0.1368	0.0187
3	0.7958	-0.1942	0.1942	0.0377
4	0.7610	-0.2290	0.2290	0.0524
5	0.8876	-0.1024	0.1024	0.0105
6	0.8114	-0.1786	0.1786	0.0319
Mean	0.8185		**MAE** = 0.1715	**MSE** = 0.0311

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 0.8185$ m
- **Variance:** Variance $= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = 0.0020$ m^2
- **SD:** SD $= \sqrt{0.0020} = 0.0447$ m

Sideseam Errors:

Table 4: Person A's Sideseam Errors - Test Case 1

Sample	**Sideseam Distance (m)**	**Error (m)**	**Absolute Error (m)**	**Squared Error (m^2)**
1	0.2694	-0.1606	0.1606	0.0258
2	0.2844	-0.1456	0.1456	0.0212
3	0.2670	-0.1630	0.1630	0.0266
4	0.2682	-0.1618	0.1618	0.0262
5	0.2828	-0.1472	0.1472	0.0217
6	0.3542	-0.0758	0.0758	0.0057
Mean	0.2877		**MAE** = 0.1423	**MSE** = 0.0212

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.2877$ m
- **Variance:** Variance $= 0.0014$ m^2
- **SD:** SD $= \sqrt{0.0014} = 0.0374$ m

Summary for Test Case 1

- **Waist Circumference:**
  - MAE: 0.1715 m
  - MSE: 0.0311 m^2
  - SD: 0.0447 m
- **Sideseam:**
  - MAE: 0.1423 m
  - MSE: 0.0212 m^2
  - SD: 0.0374 m

2.2 Test Case 2: True Depth - Predicted Keypoints

Table 5: Person A's Measurements - Test Case 2

Sample	Sideseam Distance (m)	Sideseam Distance (in)	Waistline Distance (m)	Waistline Distance (in)
1	0.3399	13.38	0.5105	20.10
2	0.3292	12.96	0.4738	18.65
3	0.3533	13.91	0.5163	20.33
4	0.3047	12.00	0.4944	19.46
5	0.3323	13.08	0.4598	18.10
6	0.4640	18.27	0.4364	17.18

Statistical Analysis for Test Case 2

- **Calculations:**

Estimated Waist Circumferences and Errors:

Table 6: Person A's Waist Circumference Errors - Test Case 2

Sample	Estimated Waist Circumference (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	1.0210	0.0310	0.0310	0.0010
2	0.9476	-0.0424	0.0424	0.0018
3	1.0326	0.0426	0.0426	0.0018
4	0.9888	-0.0012	0.0012	0.0000
5	0.9196	-0.0704	0.0704	0.0050
6	0.8728	-0.1172	0.1172	0.0137
Mean	0.9637		**MAE** = 0.0508	**MSE** = 0.0039

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 0.9637$ m
- **Variance:** Variance $= 0.0043$ m^2
- **SD:** SD $= 0.0656$ m

Sideseam Errors:

Table 7: Person A's Sideseam Errors - Test Case 2

Sample	Sideseam Distance (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.3399	-0.0901	0.0901	0.0081
2	0.3292	-0.1008	0.1008	0.0102
3	0.3533	-0.0767	0.0767	0.0059
4	0.3047	-0.1253	0.1253	0.0157
5	0.3323	-0.0977	0.0977	0.0095
6	0.4640	0.0340	0.0340	0.0012
Mean	0.3540		**MAE** = 0.0874	**MSE** = 0.0084

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.3540$ m
- **Variance:** Variance $= 0.0031$ m^2
- **SD:** SD $= 0.0557$ m

Summary for Test Case 2

- **Waist Circumference:**
  - MAE: 0.0508 m
  - MSE: 0.0039 m^2
  - SD: 0.0656 m
- **Sideseam:**
  - MAE: 0.0874 m
  - MSE: 0.0084 m^2
  - SD: 0.0557 m

2.3 Test Case 3: Predicted Depth - True Keypoints

Table 8: Person A's Measurements - Test Case 3

Sample	Left Sideseam (m)	Left (in)	Right Sideseam (m)	Right (in)	Waistline (m)	Waistline (in)
1	0.2249	8.86	0.4607	18.14	0.3504	13.79
2	0.2417	9.52	0.2776	10.93	0.3667	14.44
3	0.3676	14.47	0.2821	11.11	0.3671	14.45
4	0.2989	11.77	0.3146	12.39	0.4077	16.05
5	0.2531	9.96	0.2912	11.47	0.3770	14.84
6	0.2731	10.75	0.3323	13.08	0.3672	14.46

Statistical Analysis for Test Case 3

- **Calculations:**

Average Sideseam and Waist Circumference:

Table 9: Person A's Average Sideseam and Estimated Waist Circumference - Test Case 3

Sample	Average Sideseam (m)	Waistline (m)	Estimated Waist Circumference (m)
1	0.3428	0.3504	0.7008
2	0.2597	0.3667	0.7334
3	0.3249	0.3671	0.7342
4	0.3068	0.4077	0.8154
5	0.2722	0.3770	0.7540
6	0.3027	0.3672	0.7344

Estimated Waist Circumferences and Errors:

Table 10: Person A's Waist Circumference Errors - Test Case 3

Sample	Estimated Waist Circumference (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.7008	-0.2892	0.2892	0.0836
2	0.7334	-0.2566	0.2566	0.0659
3	0.7342	-0.2558	0.2558	0.0654
4	0.8154	-0.1746	0.1746	0.0305
5	0.7540	-0.2360	0.2360	0.0557
6	0.7344	-0.2556	0.2556	0.0653
Mean	0.7454		**MAE** = 0.2446	**MSE** = 0.0611

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 0.7454$ m
- **Variance:** Variance $= 0.0013$ m^2
- **SD:** SD $= 0.0361$ m

Sideseam Errors:

Table 11: Person A's Sideseam Errors - Test Case 3

Sample	Average Sideseam (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.3428	-0.0872	0.0872	0.0076
2	0.2597	-0.1703	0.1703	0.0290
3	0.3249	-0.1051	0.1051	0.0110
4	0.3068	-0.1232	0.1232	0.0152
5	0.2722	-0.1578	0.1578	0.0249
6	0.3027	-0.1273	0.1273	0.0162
Mean	0.3015		**MAE** = 0.1285	**MSE** = 0.0173

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.3015$ m
- **Variance:** Variance $= 0.0013$ m^2
- **SD:** SD $= 0.0361$ m

Summary for Test Case 3

- **Waist Circumference:**
  - MAE: 0.2446 m
  - MSE: 0.0611 m^2
  - SD: 0.0361 m
- **Sideseam:**
  - MAE: 0.1285 m
  - MSE: 0.0173 m^2
  - SD: 0.0361 m

2.4 Test Case 4: Predicted Depth - Predicted Keypoints

Table 12: Person A's Measurements - Test Case 4

Sample	Left Sideseam (m)	Left (in)	Right Sideseam (m)	Right (in)	Waistline (m)	Waistline (in)
1	0.3199	12.59	0.3719	14.64	0.4104	16.16
2	0.2537	9.99	0.3653	14.38	0.4130	16.26
3	0.2768	10.90	0.4256	16.75	0.4276	16.83
4	0.3198	12.59	0.3647	14.36	0.4475	17.62
5	0.2729	10.74	0.3180	12.52	0.3985	15.69
6	0.3622	14.26	0.3218	12.67	0.4092	16.11

Statistical Analysis for Test Case 4

- **Calculations:**

Average Sideseam and Waist Circumference:

Table 13: Person A's Average Sideseam and Estimated Waist Circumference - Test Case 4

Sample	Average Sideseam (m)	Waistline (m)	Estimated Waist Circumference (m)
1	0.3459	0.4104	0.8208
2	0.3095	0.4130	0.8260
3	0.3512	0.4276	0.8552
4	0.3423	0.4475	0.8950
5	0.2954	0.3985	0.7970
6	0.3420	0.4092	0.8184

Estimated Waist Circumferences and Errors:

Table 14: Person A's Waist Circumference Errors - Test Case 4

Sample	Estimated Waist Circumference (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.8208	-0.1692	0.1692	0.0286
2	0.8260	-0.1640	0.1640	0.0269
3	0.8552	-0.1348	0.1348	0.0182
4	0.8950	-0.0950	0.0950	0.0090
5	0.7970	-0.1930	0.1930	0.0372
6	0.8184	-0.1716	0.1716	0.0295
Mean	0.8354		**MAE** = 0.1546	**MSE** = 0.0249

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 0.8354$ m
- **Variance:** Variance $= 0.0015$ m^2
- **SD:** SD $= 0.0387$ m

Sideseam Errors:

Table 15: Person A's Sideseam Errors - Test Case 4

Sample	Average Sideseam (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.3459	-0.0841	0.0841	0.0071
2	0.3095	-0.1205	0.1205	0.0145
3	0.3512	-0.0788	0.0788	0.0062
4	0.3423	-0.0877	0.0877	0.0077
5	0.2954	-0.1346	0.1346	0.0181
6	0.3420	-0.0880	0.0880	0.0077
Mean	0.3311		**MAE** = 0.0990	**MSE** = 0.0102

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.3311$ m
- **Variance:** Variance $= 0.0005$ m^2
- **SD:** SD $= 0.0224$ m

Summary for Test Case 4

- **Waist Circumference:**
  - MAE: 0.1546 m
  - MSE: 0.0249 m^2
  - SD: 0.0387 m
- **Sideseam:**
  - MAE: 0.0990 m
  - MSE: 0.0102 m^2
  - SD: 0.0224 m

3 Person B's Results: Sideseam and Waistline Measurements

Assumptions and Methodology

- **Waistline Measurement Comparison:** - The ground truth waist measurement is the full waist circumference (1.23 m). - Estimated Waist Circumference = $2\times$Waistline Distance.

- **Sideseam Measurement Comparison:** - We compare the model's sideseam distances directly with the ground truth sideseam length (0.48 m).

3.1 Test Case 1: True Depth - True Keypoints

Table 16: Person B's Measurements - Test Case 1

Sample	Sideseam Distance (m)	Sideseam (in)	Waistline Distance (m)	Waistline (in)
1	0.2797	11.01	0.4878	19.20
2	0.2806	11.05	0.4728	18.62
3	0.3581	14.10	0.6174	24.31
4	0.3067	12.08	0.4788	18.85
5	0.2642	10.40	0.4538	17.86
6	0.2899	11.41	0.4912	19.34

Statistical Analysis for Test Case 1

Estimated Waist Circumference and Errors:

Table 17: Person B's Waist Circumference Errors - Test Case 1

Sample	Estimated Waist Circumference (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.9756	-0.2544	0.2544	0.0647
2	0.9456	-0.2844	0.2844	0.0809
3	1.2348	0.0048	0.0048	0.0000
4	0.9576	-0.2724	0.2724	0.0742
5	0.9076	-0.3224	0.3224	0.1039
6	0.9824	-0.2476	0.2476	0.0614
Mean	1.0006		**MAE** = 0.2310	**MSE** = 0.0642

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 1.0006$ m
- **Variance:** Variance $= \frac{1}{n-1}\sum(X_i - \overline{X})^2 = 0.0061$ m^2

- **SD:** SD $= \sqrt{0.0061} = 0.0781$ m

Sideseam Errors:

Table 18: Person B's Sideseam Errors - Test Case 1

Sample	**Sideseam Distance (m)**	**Error (m)**	**Absolute Error (m)**	**Squared Error ($\mathbf{m^2}$)**
1	0.2797	-0.2003	0.2003	0.0401
2	0.2806	-0.1994	0.1994	0.0397
3	0.3581	-0.1219	0.1219	0.0149
4	0.3067	-0.1733	0.1733	0.0300
5	0.2642	-0.2158	0.2158	0.0466
6	0.2899	-0.1901	0.1901	0.0361
Mean	0.2965		**MAE** = 0.1835	**MSE** = 0.0346

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.2965$ m
- **Variance:** Variance $= 0.0011$ m^2
- **SD:** SD $= \sqrt{0.0011} = 0.0331$ m

Summary for Test Case 1

- **Waist Circumference:**
  - MAE: 0.2310 m
  - MSE: 0.0642 m^2
  - SD: 0.0781 m
- **Sideseam:**
  - MAE: 0.1835 m
  - MSE: 0.0346 m^2
  - SD: 0.0331 m

3.2 Test Case 2: True Depth - Predicted Keypoints

Table 19: Person B's Measurements - Test Case 2

Sample	Sideseam Distance (m)	Sideseam (in)	Waistline Distance (m)	Waistline (in)
1	0.2973	11.71	0.5238	20.62
2	0.3372	13.28	0.5251	20.67
3	0.3035	11.95	0.5231	20.60
4	0.3399	13.38	0.5034	19.82
5	0.3244	12.77	0.5836	22.98
6	0.3524	13.87	0.5839	22.99

Statistical Analysis for Test Case 2

Estimated Waist Circumference and Errors:

Table 20: Person B's Waist Circumference Errors - Test Case 2

Sample	Estimated Waist Circumference (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	1.0476	-0.1824	0.1824	0.0333
2	1.0502	-0.1798	0.1798	0.0323
3	1.0462	-0.1838	0.1838	0.0338
4	1.0068	-0.2232	0.2232	0.0498
5	1.1672	-0.0628	0.0628	0.0039
6	1.1678	-0.0622	0.0622	0.0039
Mean	1.0809		**MAE** = 0.1490	**MSE** = 0.0262

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 1.0809$ m
- **Variance:** Variance $= 0.0056$ m^2
- **SD:** SD $= \sqrt{0.0056} = 0.0748$ m

Sideseam Errors:

Table 21: Person B's Sideseam Errors - Test Case 2

Sample	Sideseam Distance (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.2973	-0.1827	0.1827	0.0334
2	0.3372	-0.1428	0.1428	0.0204
3	0.3035	-0.1765	0.1765	0.0312
4	0.3399	-0.1401	0.1401	0.0196
5	0.3244	-0.1556	0.1556	0.0242
6	0.3524	-0.1276	0.1276	0.0163
Mean	0.3258		**MAE** = 0.1542	**MSE** = 0.0258

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.3258$ m
- **Variance:** $\text{Variance} = 0.0005\ \text{m}^2$
- **SD:** $\text{SD} = \sqrt{0.0005} = 0.0224$ m

Summary for Test Case 2

- **Waist Circumference:**
  - MAE: 0.1490 m
  - MSE: 0.0262 m^2
  - SD: 0.0748 m
- **Sideseam:**
  - MAE: 0.1542 m
  - MSE: 0.0258 m^2
  - SD: 0.0224 m

3.3 Test Case 3: Predicted Depth - True Keypoints

Table 22: Person B's Measurements - Test Case 3

Sample	Left Sideseam (m)	Left (in)	Right Sideseam (m)	Right (in)	Waistline (m)	Waistline (in)
1	0.2422	9.53	0.3071	12.09	0.5265	20.73
2	0.2625	10.34	0.3128	12.31	0.4988	19.64
3	0.2548	10.03	0.2888	11.37	0.5424	21.35
4	0.3339	13.15	0.3178	12.51	0.5653	22.26
5	0.2396	9.43	0.2614	10.29	0.4842	19.06
6	0.3533	13.91	0.2953	11.62	0.5179	20.39

Statistical Analysis for Test Case 3

Average Sideseam and Estimated Waist Circumference:

Table 23: Person B's Average Sideseam and Estimated Waist Circumference - Test Case 3

Sample	Average Sideseam (m)	Waistline (m)	Estimated Waist Circumference (m)
1	0.2747	0.5265	1.0530
2	0.2877	0.4988	0.9976
3	0.2718	0.5424	1.0848
4	0.3258	0.5653	1.1306
5	0.2505	0.4842	0.9684
6	0.3243	0.5179	1.0358

Waist Circumference Errors:

Table 24: Person B's Waist Circumference Errors - Test Case 3

Sample	Estimated Waist Circumference (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	1.0530	-0.1770	0.1770	0.0313
2	0.9976	-0.2324	0.2324	0.0540
3	1.0848	-0.1452	0.1452	0.0211
4	1.1306	-0.0994	0.0994	0.0099
5	0.9684	-0.2616	0.2616	0.0684
6	1.0358	-0.1942	0.1942	0.0377
Mean	1.0450		**MAE** = 0.1850	**MSE** = 0.0371

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 1.0450$ m

- **Variance:** Variance = 0.0030 m^2
- **SD:** SD = $\sqrt{0.0030}$ = 0.0548 m

Sideseam Errors:

Table 25: Person B's Sideseam Errors - Test Case 3

Sample	Average Sideseam (m)	Error (m)	Absolute Error (m)	Squared Error (m^2)
1	0.2747	-0.2053	0.2053	0.0421
2	0.2877	-0.1923	0.1923	0.0370
3	0.2718	-0.2082	0.2082	0.0433
4	0.3258	-0.1542	0.1542	0.0238
5	0.2505	-0.2295	0.2295	0.0527
6	0.3243	-0.1557	0.1557	0.0243
Mean	0.2891		**MAE** = 0.1909	**MSE** = 0.0372

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.2891$ m
- **Variance:** Variance = 0.0041 m^2
- **SD:** SD = 0.0639 m

Summary for Test Case 3

- **Waist Circumference:**
  - MAE: 0.1850 m
  - MSE: 0.0371 m^2
  - SD: 0.0548 m
- **Sideseam:**
  - MAE: 0.1909 m
  - MSE: 0.0372 m^2
  - SD: 0.0639 m

3.4 Test Case 4: Predicted Depth - Predicted Keypoints

Table 26: Person B's Measurements - Test Case 4

Sample	Left Sideseam (m)	Left (in)	Right Sideseam (m)	Right (in)	Waistline (m)	Waistline (in)
1	0.3439	13.54	0.3963	15.60	0.5120	20.16
2	0.2912	11.46	0.3740	14.73	0.4913	19.34
3	0.2883	11.35	0.2669	10.51	0.4922	19.38
4	0.2955	11.63	0.3577	14.08	0.5764	22.69
5	0.3730	14.69	0.3347	13.18	0.5452	21.46
6	0.2737	10.78	0.4033	15.88	0.5343	21.03

Statistical Analysis for Test Case 4

Average Sideseam and Estimated Waist Circumference:

Table 27: Person B's Average Sideseam and Estimated Waist Circumference - Test Case 4

Sample	Average Sideseam (m)	Waistline (m)	Estimated Waist Circumference (m)
1	0.3701	0.5120	1.0240
2	0.3326	0.4913	0.9826
3	0.2776	0.4922	0.9844
4	0.3266	0.5764	1.1528
5	0.3539	0.5452	1.0904
6	0.3385	0.5343	1.0686

Waist Circumference Errors:

Table 28: Person B's Waist Circumference Errors - Test Case 4

Sample	Estimated Waist Circumference (m)	Error (m)	Abs. Error (m)	Squared Error (m^2)
1	1.0240	-0.2060	0.2060	0.0424
2	0.9826	-0.2474	0.2474	0.0612
3	0.9844	-0.2456	0.2456	0.0603
4	1.1528	-0.0772	0.0772	0.0060
5	1.0904	-0.1396	0.1396	0.0195
6	1.0686	-0.1614	0.1614	0.0260
Mean	1.0505		**MAE** = 0.1795	**MSE** = 0.0359

Standard Deviation (SD) Calculation:

- **Mean Estimated Waist Circumference:** $\overline{X} = 1.0505$ m

- **Variance:** Variance $= 0.0051\ \text{m}^2$
- **SD:** SD $= \sqrt{0.0051} = 0.0714$ m

Sideseam Errors:

Table 29: Person B's Sideseam Errors - Test Case 4

Sample	Average Sideseam (m)	Error (m)	Abs. Error (m)	Squared Error (m^2)
1	0.3701	-0.1099	0.1099	0.0121
2	0.3326	-0.1474	0.1474	0.0217
3	0.2776	-0.2024	0.2024	0.0410
4	0.3266	-0.1534	0.1534	0.0235
5	0.3539	-0.1261	0.1261	0.0159
6	0.3385	-0.1415	0.1415	0.0200
Mean	0.3332		**MAE** = 0.1468	**MSE** = 0.0224

Standard Deviation (SD) Calculation:

- **Mean Sideseam Distance:** $\overline{X} = 0.3332$ m
- **Variance:** Variance $= 0.0010\ \text{m}^2$
- **SD:** SD $= 0.0316$ m

Summary for Test Case 4

- **Waist Circumference:**
  - MAE: 0.1795 m
  - MSE: 0.0359 m^2
  - SD: 0.0714 m
- **Sideseam:**
  - MAE: 0.1468 m
  - MSE: 0.0224 m^2
  - SD: 0.0316 m