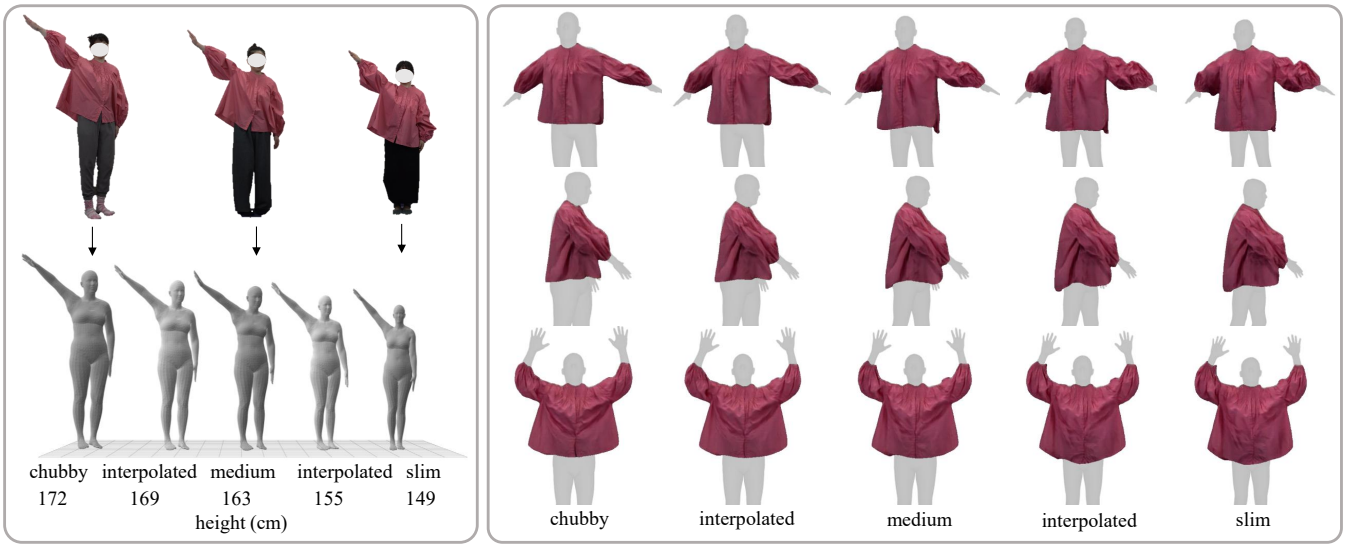
Capturing Size-Aware Draping via Per-Garment 2D Try-On

Jingyuan LIU
The University of Tokyo
Tokyo, Japan
jliucb@connect.ust.hk

Zaiqiang Wu
The University of Tokyo
Tokyo, Japan

Yechen Li
The University of Tokyo
Tokyo, Japan

Takeo Igarashi
The University of Tokyo
Tokyo, Japan
takeo@acm.org



(a) capture (b) synthesize

Figure 1: We propose a size-aware virtual try-on approach that simulates the degree of tight or loose draping on specific body sizes directly in 2D, enabling efficient interactive applications. Our per-garment method follows a capture-and-synthesis pipeline: (a) we first capture garment appearances from video under various poses and body sizes; (b) we then train a generative neural network to synthesize garment images for a given body size. It also generalizes to unseen interpolated body shapes.

1 Introduction

In virtual try-on applications, accurately showing how garments of different sizes drape on a subject-specific body shape is crucial for informed decision-making. While 3D garment simulation provides the most accurate results, it requires detailed 3D garment models and realistic physical parameters, and its computational cost limits interactive use. Recent advances in generative AI offer an efficient 2D alternative, motivating our exploration of size-aware virtual try-on using generative neural networks.

Extensive existing 2D virtual try-on methods focus on aligning a clothing product image to a parsed region of a subject image to generate a try-on preview. Among them, two notable methods have attempted to incorporate garment size. Specifically, COTTON [Chen et al. 2023] uses the shoulder-to-torso length ratio as a body size measure, influencing alignment between the garment image and body landmarks. SiCo [Chen et al. 2025] employs discrete size labels (e.g., XS, L) to qualitatively adjust the scale of the clothing segmentation mask. Although these methods produce visually appealing results with some variation in garment shape, they inherit a core limitation of 2D virtual try-on: the deformation of garments relies on predefined (e.g., Thin Plate Spline) or learned transformations from a broad garment dataset, which may not accurately reflect the real physical properties of a specific garment. As a result, they often fail to preserve the physical identity of a garment, such as realistic shape deformations and wrinkles. This highlights the need to capture garment-specific properties to simulate the authentic draping effects in synthesis.

To address this, a distinct line of work—per-garment 2D virtual try-on—focuses on first capturing a garment's appearance under various conditions, then synthesizing it under new video subject conditions. Representative methods [Chong et al. 2021; Wu et al. 2024] condition synthesis on body pose and viewpoint. We build upon this framework and extend it by incorporating body size as an additional conditioning factor. This is achieved through a dedicated neural network architecture and the construction of size-aware, per-garment image datasets. Experiments demonstrate that our method efficiently and realistically recovers garment appearance while generalizing to unseen body shapes.

2 Method

Our approach follows a capture-and-synthesis mechanism. During the capture (Figure 1(a)), we build a garment-specific dataset by recording multiple human subjects—varying in body shape—wearing the same garment while performing actions. We formulate clothes

SA Posters '25, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2134-2/2025/12
https://doi.org/10.1145/3757374.3771424

image synthesis as a conditional generation task: $I_{cloth} = H_s(p, m, \sigma)$, where p is the body pose feature, m is the body size feature, and σ is the viewpoint. H_s is a conditional generative neural network trained specifically for garment size s, (e.g., S, M, L). Since each garment size represents a distinct physical configuration, we train a separate model H_s for each. During inference (Figure 1(b)), the system extracts (p, m, σ) from an image of the subject and synthesizes the garment appearance. Previewing different garment sizes is achieved by switching between the corresponding models H_s.

To encode pose and viewpoint, we adopt the representation proposed in [Wu et al. 2024]. For each input frame, we first reconstruct the subject's pose and body shape using an off-the-shelf SMPL estimator. The reconstructed SMPL mesh is then textured and rendered into a 2D pose map, from which convolutional features are extracted (Figure 2(b)), forming the representation of p and σ.

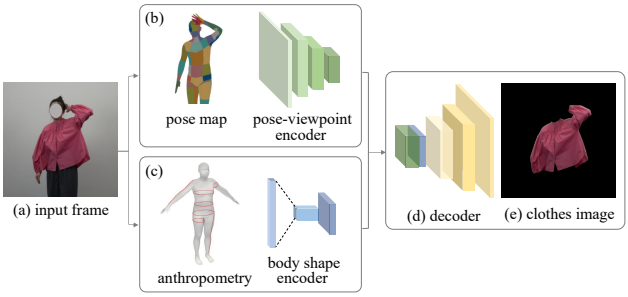


Figure 2: An overview of the network architecture.

We introduce an additional shape encoding branch (Figure 2(c)) to condition on body size. Inspired by tailoring practices, body size is represented using anthropometric measurements (e.g., shoulder width, waist circumference) derived from the reconstructed SMPL mesh by cross-sectioning it with predefined planes in A-pose. To address the scale ambiguity introduced by unknown subject–camera distance, we explicitly collect the subject's physical height and compute its ratio to the reconstructed SMPL model height, normalizing all other measurements accordingly. The resulting measurements form a 1D body size feature vector, which is subsequently encoded into a 2D feature map m.

The pose-viewpoint features and the shape features are concatenated and passed to the decoder to generate the final garment image (Figure 2(d-e)). The network is trained using a combination of GAN loss, feature matching loss, and perceptual (VGG) loss, following [Wu et al. 2024].

3 Experiments

To evaluate our approach, we collected a training set for each garment using the setup in [Wu et al. 2025]. We invited three human subjects with chubby, medium, and slim body types. See the supplementary video for the collection process. Each dataset contains approximately 24,000 samples. Experiments were conducted on a PC running Ubuntu 20.04 with an NVIDIA RTX 3080Ti GPU. Training for 100 epochs took about 4 days per model; inference runs at about 9 FPS.

We tested the trained network on novel try-on poses. Qualitative results (Figure 3) show that our method realistically captures the interaction between body shape and garment size—for example, generating looser sleeves and longer hems for slim bodies compared to chubby ones. It also generalizes well to interpolated body shapes not seen during training (Figure 1(b)). Please see the supplementary video for more results.

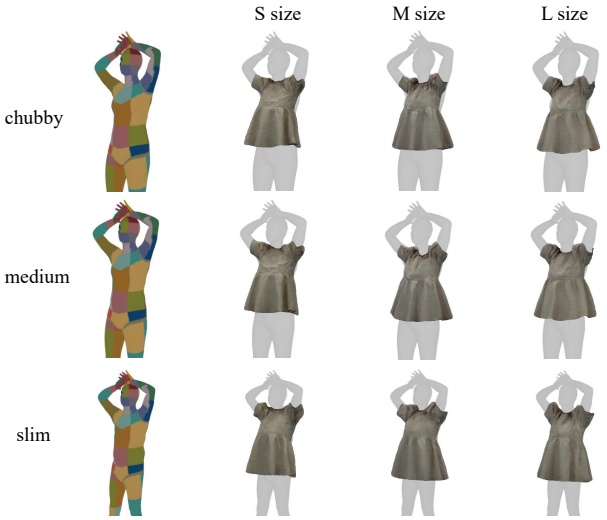


Figure 3: Synthesized results of three garment sizes on three different body shapes.

We compare our method against two size-aware, image-based baselines: COTTON [Chen et al. 2023] and SiCo [Chen et al. 2025]. As shown in Figure 4, our method produces more realistic draping and fit compared to these baselines. Notably, our synthesized images more closely match the ground-truth appearance, emphasizing the importance of explicit garment-specific data capture.

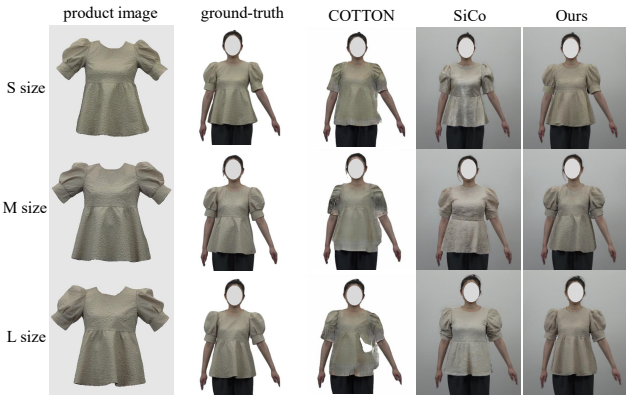


Figure 4: Comparison with COTTON and SiCo.

4 Conclusions

We present a method to capture and synthesize garment appearance across varying body shapes, enabling size-aware VTON. Evaluated on a toy dataset with three body shapes, the approach delivers promising results. In future work, we will curate a larger dataset with more diverse body shapes to strengthen model generalization.

References

Chieh-Yun Chen, Yi-Chung Chen, Hong-Han Shuai, and Wen-Huang Cheng. 2023. 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*. 7513–7522.

Sherry X Chen, Alex Christopher Lim, Yimeng Liu, Pradeep Sen, and Misha Sra. 2025. SiCo: An Interactive Size-Controllable Virtual Try-On Approach for Informed Decision-Making. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 1815–1825.

Toby Chong, I-Chao Shen, Nobuyuki Umetani, and Takeo Igarashi. 2021. Per garment capture and synthesis for real-time virtual try-on. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 457–469.

Zaiqiang Wu, Yechen Li, Jingyuan Liu, Yuki Shibata, Takayuki Hori, I Shen, Takeo Igarashi, et al. 2025. Low-Barrier Dataset Collection with Real Human Body for Interactive Per-Garment Virtual Try-On. *arXiv preprint arXiv:2506.10468* (2025).

Zaiqiang Wu, Jingyuan Liu, Toby Chong, I-Chao Shen, and Takeo Igarashi. 2024. Virtual Measurement Garment for Per-Garment Virtual Try-On. In *Proceedings of the 50th Graphics Interface Conference*. 1–10.