FIT: A Large-Scale Dataset for Fit-Aware Virtual Try-On

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 980

ACM Reference Format:
Anonymous Author(s). 2026. FIT: A Large-Scale Dataset for Fit-Aware Virtual Try-On. 1, 1 (February 2026), 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

A Additional FIT Dataset Details

In this section, we provide additional details and statistics about the FIT dataset.

A.1 Size Categorization

In our figures, we frequently abbreviate the full person-garment measurements $[m_p, m_g]$ with coarse size labels for the person and garment such as XS, L, XL. We determine these size labels based on average measurement ranges as shown in Table 3. Note that the grouping of **size measurements** into **coarse size labels** (e.g. XS, L, XL) are only used for visualization and grouping purposes, not Fit-VTO training or evaluation.

A.2 Garment Fit Distribution

Our FIT dataset covers a diverse range of fit scenarios. In Figure 4, we plot the distribution of person/garment size pairings as a histogram, showing that every reasonable fit scenario is represented, from very tight (e.g. size "XL" person wearing a size"M" garment) to very loose (e.g. size "XS" person wearing a size "2XL" garment). Implausible fit pairings where the garment is more than 3 sizes smaller than the person (e.g. size "3XL" person wearing a size"XS" garment) are not included.

A.3 Measurement Statistics

In Table 1 and Table 2, we report the minimum, mean, maximum, and standard deviations of the measurements for our body and garment meshes in the FIT dataset, respectively. Our dataset covers a wide range of body shapes and garment sizes.

Table 1. Body size statistics. We report the min, mean, max, and standard deviation of our garment measurements in cm.

Measurement	**Men's** (min, mean, max, std)	**Women's** (min, mean, max, std)
Bust	(87, 101, 141, 10)	(83, 100, 136, 13)
Height	(155, 174, 194, 8.0)	(151, 170, 196, 9.0)
Hips	(88, 101, 125, 6.0)	(89, 104, 127, 10)
Waist	(70, 86, 141, 13)	(61, 85, 130, 17)

ACM XXXX-XXXX/2026/2-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Table 2. Garment size statistics. We report the min, mean, max, and standard deviation of our garment measurements in cm.

Measurement	**Men's** (min, mean, max, std)	**Women's** (min, mean, max, std)
Width	(77, 112, 169, 16)	(75, 110, 169, 16)
Length	(29, 53, 76, 8.0)	(29, 51, 76, 7.5)
Sleeve Length	(0.0, 30, 79, 17)	(0, 29, 79, 17)

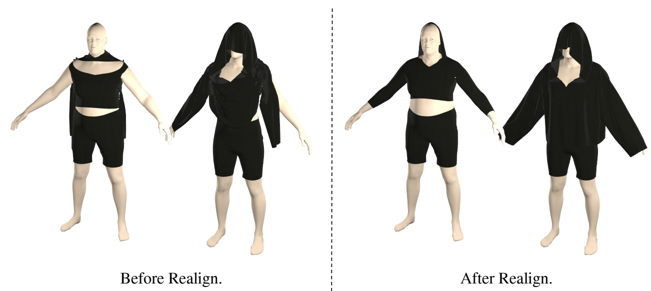

Fig. 1. During cross-body draping, the initial boxmeshes are often misaligned with the target human models, causing draping failures (left). We explicitly realign the boxmesh to ensure successful simulation (right).

A.4 Additional Dataset Examples

We present additional examples of try-on triplet data in our dataset in Figure 2.

B Additional Details on Data Generation Pipeline

B.1 Cross Draping vs. Linear Size Change

The standard GarmentCode pipeline computes sewing patterns based on a design template and target body parameters, yielding a 3D garment that is well-fitted to the wearer. To generate ill-fitting examples (e.g., oversized or undersized), a naive baseline would be to linearly scale the garment parameters. However, this fails to capture real-world sizing dynamics, as garment grading rules are non-linear and distinct from simple geometric scaling. To address this, we propose a cross-draping strategy. Instead of manipulating the mesh directly, we instantiate a separate "source" body in a different size, generate a pattern fitted to that body, and then drape the resulting garment onto the original target body. This process simulates the physical reality of a person wearing a garment designed for someone else, resulting in significantly more natural and realistic ill-fitting dynamics compared to simple linear scaling.

B.2 Boxmesh Realignment

Cross-draping a sewing pattern onto a target body mesh of a different size creates misalignments between the boxmesh panels and target body parts, which can lead to draping errors. We implement

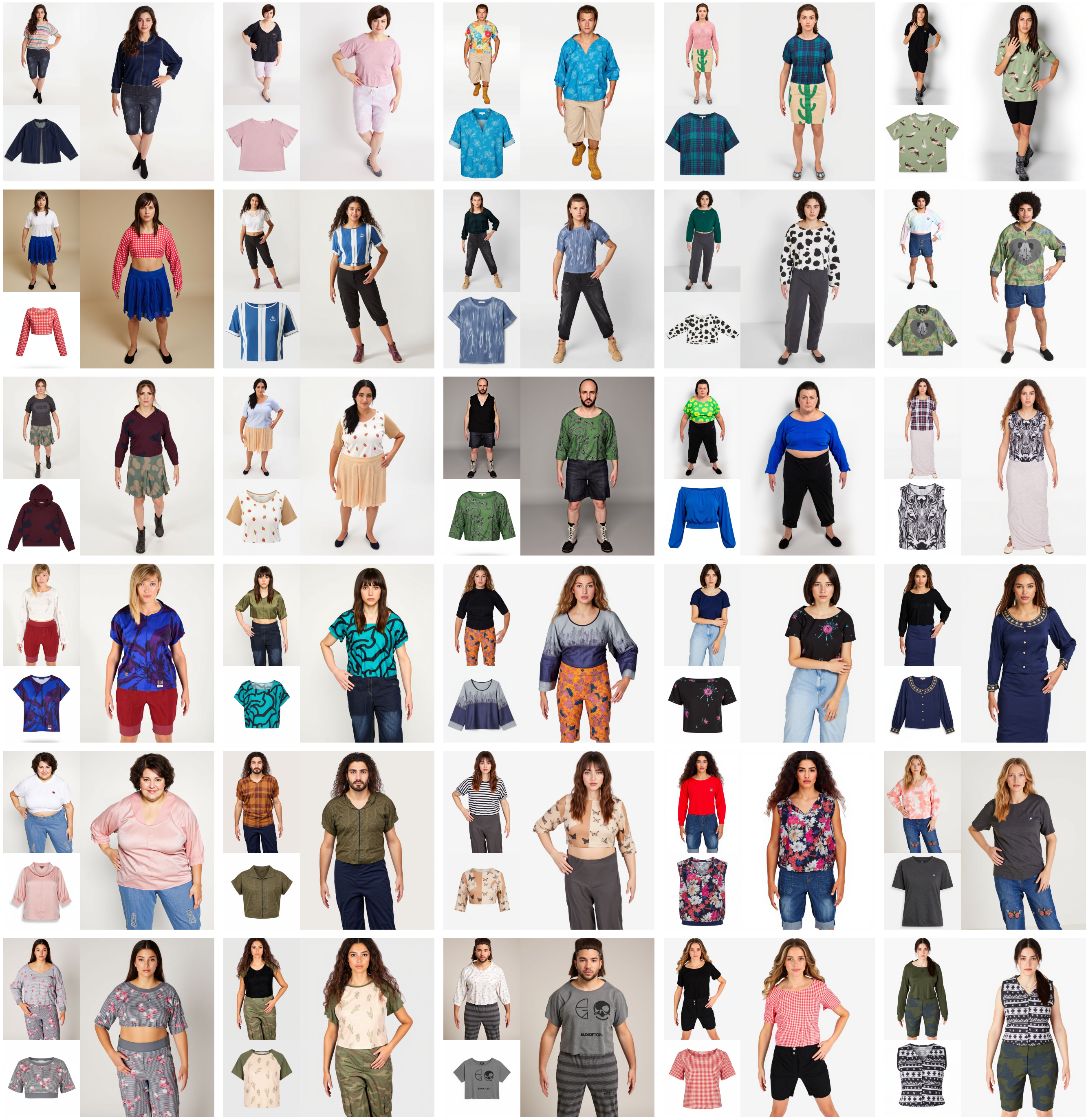

Fig. 2. **Additional dataset examples.** In each example, we show the paired person image (top left), garment image (lower left), and the target try-on image (right).

Table 3. Body size categorization statistics. We report the min and max for each body measurements and size label in cm.

	Bust		Waist		Hips	
Size	**Men's** (min, max)	**Women's** (min, max)	**Men's** (min, max)	**Women's** (min, max)	**Men's** (min, max)	**Women's** (min, max)
XS	(86, 91)	(79, 84)	(71, 76)	(58, 64)	(91, 96)	(84, 89)
S	(91, 96)	(86, 89)	(76, 81)	(66, 67)	(96, 101)	(91, 94)
M	(96, 101)	(90, 95)	(81, 86)	(71, 75)	(101, 106)	(97, 102)
L	(101, 106)	(96, 104)	(86, 91)	(86, 91)	(106, 111)	(106, 111)
XL	(106, 117)	(105, 116)	(91, 103)	(85, 97)	(111, 120)	(112, 121)
2XL	(111, 127)	(112, 125)	(96, 115)	(91, 105)	(120, 134)	(120, 130)
3XL	(127, 147)	(117, 135)	(115, 137)	(107, 127)	(134, 145)	(125, 137)

boxmesh realignment (Section 3.2 in main) as a critical step for successful cross-body draping. See Figure 1 for a visual comparison with and without boxmesh alignment.

To align a given sewing pattern p_{tgt}(might be ill-fit to size s_p body) with a target body mesh of size s_p, we use a different, well-fitted (i.e. generated on size s_p body) sewing pattern p_{ref} as a reference. We then align the panels of the p_{tgt} to the spatial locations of p_{ref}. This ensures that p_{tgt} is aligned to the target body mesh. Additionally, we observed that significant discrepancies between the human model's arm angle and the initialized sleeve panel angle can cause arm-sleeve penetrations. To mitigate this, we adjust the sleeve angle to match the arm angle prior to simulation.

B.3 Retexturing Model Training Data

We train our retexturing model on a dataset of 50k real-world person images. We use the person images in VITON-HD and additionally scraped online images featuring modeling posing in front of the camera with a studio background. We use Sapiens [Khirodkar et al. 2024] to estimate normal and segmentation maps and Gemini [Google 2025a] to generate prompts describing the garment textures and designs. We enforce a structured prompt format containing two sentences (one per garment piece). This facilitates inference for paired generation: since only the top garment is swapped, we update the first sentence corresponding to the top, while the second sentence remains frozen to preserve the bottom garment.

B.4 Reposing

Since GarmentCode simulation produces results exclusively in a static A-pose, we employ a customized reposing pipeline to repose the 3D simulated meshes, thereby expanding the dataset's diversity and improve model generalization. In total, we sample from 528 distinct target poses and repose each sample into a randomly chosen pose from the pool, prioritizing casual stances commonly encountered in real-world try-on scenarios.

C Preprocessing Details for Online Fashion Image Dataset

To generate paired-person images I_p for the online fashion image dataset, we leverage Nano Banana Pro [Google 2025b] with the prompt in Section G.4. To train our ablation without FIT data, we use Gemini [Google 2025a] to estimate coarse measurements using the prompt in Section G.5, since online fashion images do not include size annotations. We refer to these measurements as "coarse", because they are merely rough estimates due to scale ambiguity from single image.

D Additional Resizing Results

To show how Fit-VTO generalizes to real-world images, we provide additional resizing results on real-world person images in Figure 5. In these examples, the person and person measurements are captured from real human subjects, and the garment layflat image and measurements are randomly chosen from the FIT test dataset.

E Failure Cases

We show qualitative examples of the limitations of our method in Figure 3. These include a limited ability to represent varying degrees of garment tightness in GarmentCode [Korosteleva and Sorkine-Hornung 2023], which we leave to future work. Another limitation is that garment measurements are often correlated in our FIT data (e.g. larger width correlates positively with larger length). As a result, with our Fit-VTO model, changing one measurement may lead to an unintentional change in another dimension.

F Comparisons to State-of-the-Art

F.1 VITON-HD Preprocessing

Due to the lack of paired data, we generated pseudo paired-person images I_p for every image in the VITON-HD [Choi et al. 2021] using Nano Banana Pro (see Section G.4 for prompts). When running our Fit-VTO method, we set the each person and garment measurement to the null value (-1), same as the dropout value used during training.

F.2 Implementation Details

Any2AnyTryon [Guo et al. 2025]: We used the model and code released from the official implementation. For all evaluations, we used the "dev_lora_any2any_multi" checkpoint.

COTTON [Chen et al. 2023]: The officially released code and checkpoint trained on COTTON dataset was used. For evaluation on the VITON-HD test dataset, the default try-on mode was used. When running on FIT test dataset, the scaling parameter was computed as r = length/bust.

Nano Banana Pro [Google 2025b]: For comparisons to Nano Banana Pro [Google 2025b], we input the paired-person image I_p,

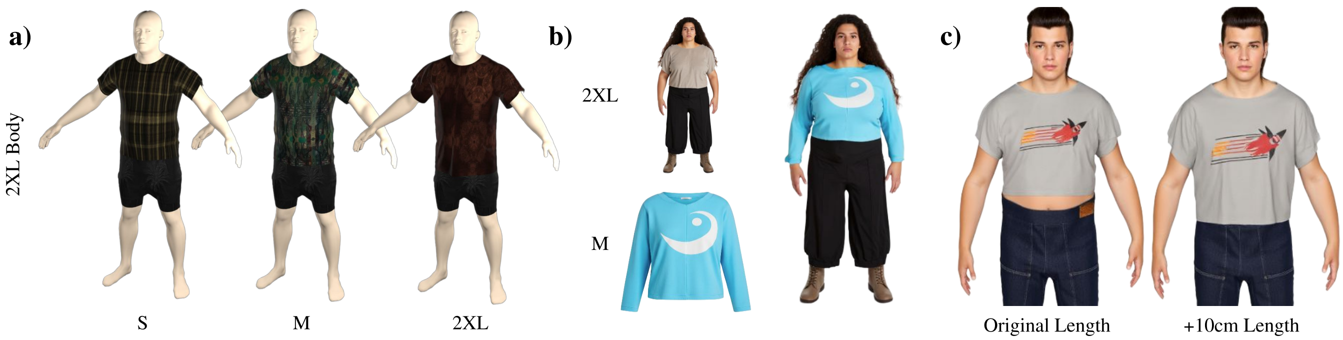

Fig. 3. Failure cases. (a) GarmentCode [Korosteleva and Sorkine-Hornung 2023] simulation does not model varying degrees of tightness well, leading to similar-looking fit for all garment sizes smaller than the body size. (b) As a result of (a), it is difficult to tell the level of tightness in FIT try-on images. (c) Due to the correlations in measurements across sizes, adjustments to individual measurements may also lead to undesired changes in other dimensions. In this example, increasing garment length also increases garment width.

layflat garment image I_g, person-garment measurements m, and the prompt:

> Edit $image_0$ so that the person wears garment in $image_1$ with size of person and garment described as {measurement description}.

IDM-VTON [Choi et al. 2024]: We used the model and code released from the official implementation. For VITON-HD, the agnostic masks from the original data release were used. For FIT dataset, agnostic masks were computed from IDM-VTON preprocessing code. All hyper-parameters(e.g. number of diffusion steps) are set to be the recommended value from official release.

G LLM and VLM Prompts

In the follow sections, we provide the exact prompts used for all calls to LLM (Gemini [Google 2025a]) and VLM (Nano Banana Pro [Google 2025b]) models in this paper.

G.1 Head & Shoes Generation

> Change the head to make it look photorealistic. Add realistic <hair style> hair, but the hair should always be behind the shoulder and never at the front. Add <shoe type> if feet are visible. Make sure that everything else stays identical, including the human pose, garment shape, size, design and position.

G.2 Prompt Generation

> Describe the garment in the image in two sentences. The first sentence should describe the top garment, and the second sentence should describe the bottom garment. Note that the input image is an illustration of the garment type, style and size - please ignore its existing texture. Please come up with some new description of the texture, logo and design. Add pocket, zipper, button, and other garment details if appropriate. Keep everything under 50 words.

G.3 Garment Try-Off

> Create an in-shop product image of the top garment only against a plain white background.

G.4 Paired Person Image Generation

> Generate a new image where the upper garment is changed, and keep everything else exactly the same, including the bottom garment, face, human pose, position etc.

G.5 Coarse Measurement Estimate

> Estimate the height, bust, hips and waist of the human in cm. Also estimate the corresponding bust, length (shoulder strap to bottom hem) and sleeve length of the top garment this person is wearing. Formulate the results like this: Human Height: xx cm, Human Bust: xx cm, Human Hips: xx cm, Human Waist: xx cm, Garment Length: xx cm, Garment Sleeve Length: xx cm, Garment Bust: xx cm.
> Please provide precise measurements in cm, not a range. Please strictly follow the format exactly. Do not add anything else to the response.

G.6 Quality Assurance (QA)

We leverage our LLM to filter out draping errors in I_s that expose either person bust or groin area. We use the following two prompts to detect such errors:

> Does the garment cover the person's chest? If so, return 'pass'. If not, return 'fail'.

> Does the image contain a bottom garment (skirt, pants, underwear, boxers, leggings, or shorts) that covers the person's groin area? If so, return 'pass'. If not, return 'fail'.

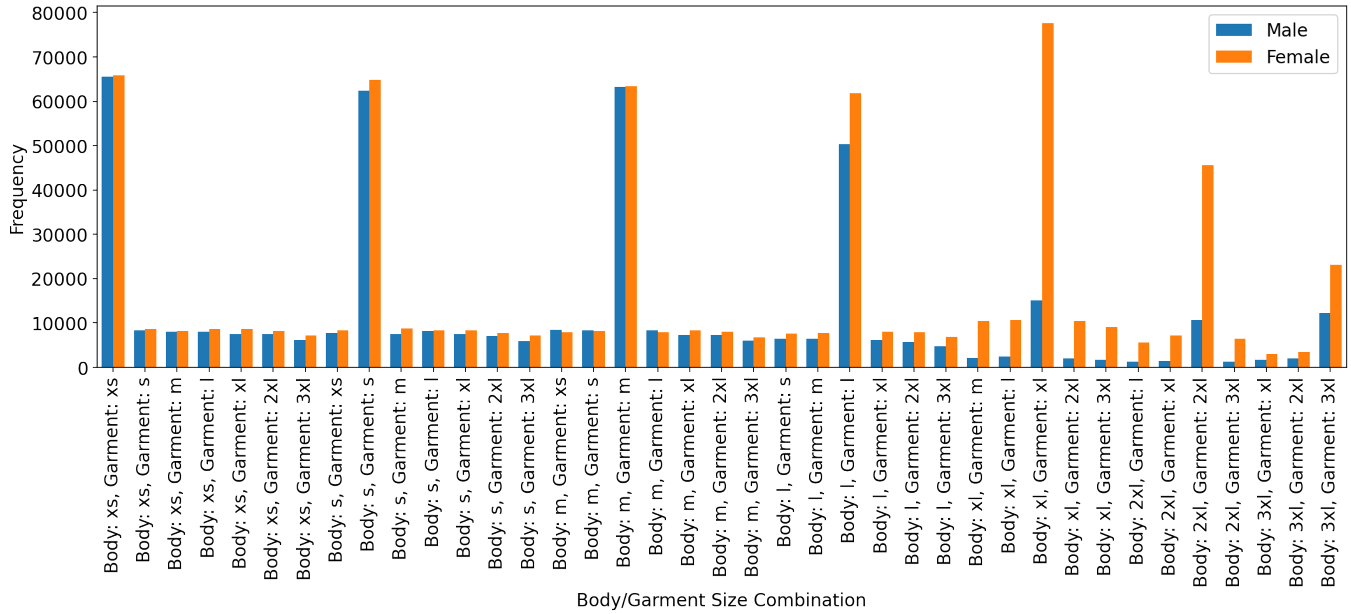

Fig. 4. Garment fit distribution. We plot the frequency of each (body size, garment size) pairing in our dataset according to the size classification in Table 3.

Fig. 5. Real-world resizing results. We show Fit-VTO try-on performance on real-world person images using varying garment sizes. Fit-VTO realistically shrinks and grows the garment fit according to uniform adjustments to the garment measurements–length, width, and sleeve length–with respect to their original values (1.0x). The size label to the left of each example corresponds to the person's body size. Since real-world garment images with precise measurements are difficult to acquire, we use our synthetic garment images and measurements.

H Usage of LLM's

In addition to using an LLM [Google 2025a] as described in Sections A, F, and G, we leveraged an LLM to improve the grammar and clarity of our writing.

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.

Seunghwan Choi, Sunghyun Park, Minsoo Lee, and Jaegul Choo. 2021. VITON-HD: High-Resolution Virtual Try-On via Misalignment-Aware Normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yisol Choi, Sangkyung Kwak, Kyungmin Lee, Hyungwon Choi, and Jinwoo Shin. 2024. Improving Diffusion Models for Authentic Virtual Try-on in the Wild. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Google. 2025a. Gemini. https://gemini.google.com/.

Google. 2025b. Gemini 3 Pro Image. https://gemini.google.com/.

Hailong Guo, Bohan Zeng, Yiren Song, Wentao Zhang, Chuan Zhang, and Jiaming Liu. 2025. Any2AnyTryon: Leveraging Adaptive Position Embeddings for Versatile Virtual Clothing Tasks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Rawal Khirodkar, Timur Bagautdinov, Julieta Martinez, Su Zhaoen, Austin James, Peter Selednik, Stuart Anderson, and Shunsuke Saito. 2024. Sapiens: Foundation for human vision models. In *European Conference on Computer Vision*. Springer, 206–228.

Maria Korosteleva and Olga Sorkine-Hornung. 2023. GarmentCode: Programming Parametric Sewing Patterns. *ACM Transactions on Graphics (TOG)* 42, 6, Article 197 (2023). doi:10.1145/3618351