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

JOHANNA KARRAS*, University of Washington, USA and Google Research, USA
YUANHAO WANG*, University of Washington, USA and Google Research, USA
YINGWEI LI, Google Research, USA
IRA KEMELMACHER-SHLIZERMAN, University of Washington, USA and Google Research, USA


Fig. 1. **The FIT Dataset.** We present FIT, a dataset and benchmark designed for *fit-aware* virtual try-on, featuring diverse garment fits (e.g., tight, loose) and precise size annotations. *Left:* Sample dataset triplets showing the conditioning garment image (*top*), the conditioning person image (*middle*), and the target try-on image (*bottom*). *Right:* Visualization of the corresponding person and garment measurement annotations. Backgrounds are removed for clarity.

Given a person and a garment image, virtual try-on (VTO) aims to synthesize a realistic image of the person wearing the garment, while preserving their original pose and identity. Although recent VTO methods excel at visualizing garment appearance, they largely overlook a crucial aspect of the try-on experience: the accuracy of garment fit – for example, depicting how an extra-large shirt looks on an extra-small person. A key obstacle is the absence of datasets that provide precise garment and body size information, particularly for "ill-fit" cases, where garments are significantly too large or too small. Consequently, current VTO methods default to generating well-fitted results regardless of the garment or person size.

In this paper, we take the first steps towards solving this open problem. We introduce FIT (**F**it-**I**nclusive **T**ry-on), a large-scale VTO dataset comprising over **1.13M** try-on image triplets accompanied by precise body and garment measurements. We overcome the challenges of data collection via a scalable synthetic strategy: (1) We programmatically generate 3D garments using GarmentCode [Korosteleva and Sorkine-Hornung 2023] and drape them via physics simulation to capture realistic garment fit. (2) We employ a novel re-texturing framework to transform synthetic renderings into photo-realistic images while strictly preserving geometry. (3) We introduce person identity preservation into our re-texturing model to generate paired person images (same person, different garments) for supervised training. Finally, we leverage our FIT dataset to train a baseline fit-aware virtual try-on model. Our data and results set the new state-of-the-art for fit-aware virtual try-on, as well as offer a robust benchmark for future research. We will make all data and code publicly available.

CCS Concepts: • **Computing methodologies** → *Computer vision.*

Additional Key Words and Phrases: Virtual Try-On, diffusion model, sim2real

*Both authors contributed equally to this research.

Authors' Contact Information: Johanna Karras, Paul G. Allen School for Computer Science and Engineering, University of Washington, Seattle, WA, USA, and Google Research, Seattle, CA, USA, jskarras@cs.washington.edu; Yuanhao Wang, Paul G. Allen School for Computer Science and Engineering, University of Washington, Seattle, WA, USA, and Google Research, Seattle, CA, USA, yuanhaowang@cs.washington.edu; Yingwei Li, , Google Research, Mountain View, CA, USA, yingweili@google.com; Ira Kemelmacher-Shlizerman, Paul G. Allen School for Computer Science and Engineering, University of Washington, Seattle, WA, USA, and Google Research, Seattle, CA, USA, kemelmi@google.com.


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

**ACM Reference Format:**
Johanna Karras, Yuanhao Wang, Yingwei Li, and Ira Kemelmacher-Shlizerman. 2026. FIT: A Large-Scale Dataset for Fit-Aware Virtual Try-On. 1, 1 (February 2026), 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 Introduction

The rising popularity of online shopping and social media has increased the demand for virtual try-on (VTO) systems. Driven by advances in generative models, recent VTO works [Chong et al. 2024; Guo et al. 2025; Zhu et al. 2024] have achieved remarkable progress in synthesizing photorealistic try-on images. However, they often merely transfer garment *appearance* onto a person, neglecting to take into account the person or garment sizes. As such, current VTO methods fail to address a fundamental question for any user: *"How will this garment actually fit me?"* This severely limits the accuracy and reliability of existing VTO tools to simulate a real-life try-on experience. Furthermore, it prevents users from experimenting with different sizes to achieve a desired fitted or oversized look. Consequently, there is significant commercial and research interest in developing a fit-aware VTO method.

Fit-aware try-on remains challenging due to the scarcity of real-world data annotated with precise person and garment measurements. Most existing VTO datasets [Bertiche et al. 2020; Choi et al. 2021; Cui et al. 2023; Ge et al. 2019; Han et al. 2018; Liu et al. 2023, 2016; Morelli et al. 2022; Patel et al. 2020; Zhu et al. 2020; Zou et al. 2023] are curated by scraping catalog images from online retailers, which inherently lack "ill-fit" examples, i.e. the garment is too large or too small. Moreover, while some retailers provide size metadata, these annotations are often non structured and difficult to process at scale. Synthetic 3D garments created by artists offer an alternative, but this data suffers from limited scale and realism.

To fill this gap, we introduce FIT (**F**it-**I**nclusive **T**ry-on), the first large-scale, size-aware VTO benchmark explicitly designed to capture diverse fit scenarios. By pivoting to a synthetic data generation pipeline (GarmentCode [Korosteleva and Sorkine-Hornung 2023]), we overcome the limitations of real-world data collection. We procedurally create 3D garments with exact ground-truth measurements and simulate their drape onto a wide range of parametric bodies. This approach ensures not only size measurements, but also details like wrinkles, stretch, and garment coverage, are physically accurate. To close the domain gap between synthetic and real images, we employ a novel re-texturing pipeline designed to generate photorealistic textures for the synthetic renderings, while ensuring that the garment fit and body shape are preserved. To this end, we fine-tune a foundational image generation model, Flux.1-dev[Black Forest Labs 2024], to generate realistic person images from the synthetic normal maps and text-based garment descriptions.

Another critical bottleneck in VTO research is the lack of paired training data (identical subject and pose, different garments). Consequently, existing methods [Chong et al. 2024; Kim et al. 2024, 2025; Xu et al. 2025; Zhu et al. 2024, 2023] are forced to formulate VTO as a self-supervised reconstruction task, which limits real-world applications, or rely on synthesized pseudo triplets [Du et al. 2023; Guo et al. 2025; Zhang et al. 2025], which suffer from inaccurate masking, identity loss, and size leakage. In contrast, our synthetic pipeline offers the unique advantage of controllability. We can simulate the same 3D subject in the same pose wearing multiple distinct garments, thereby generating ground-truth paired person data. Building on this insight, we further propose a novel framework for paired person image generation that ensures accurate 3D grounding and identity preservation.

Our dataset contains 1.13M training and 1K test samples of both men's and women's garments. Each sample consists of a target try-on image, layflat garment image, a paired person image, as well as person and garment measurements. Our target try-on images cover diverse fit scenarios, including extreme ill-fits (e.g. a size 3XL draped onto a size XS person). By fine-tuning Flux.1-dev[Black Forest Labs 2024] with our custom dataset and a custom measurement encoder, we demonstrate a baseline fit-aware VTO model that accurately showcases garment fit.

To summarize, we present the following contributions:

(1) We introduce **FIT**, the first large-scale dataset and benchmark explicitly designed for fit-aware virtual try-on, featuring precise metric annotations and diverse fit scenarios.
(2) We develop a scalable synthetic data generation pipeline that leverages physics simulation and generative re-texturing to produce photorealistic try-on triplets with 3D grounding.
(3) We demonstrate a novel, fit-aware virtual try-on model (**Fit-VTO**) that incorporates person and garment measurements to visualize not only garment appearance, but also accurate garment fit.

## 2 Related Works

### 2.1 Virtual Try-On Datasets

A primary bottleneck for fit-aware virtual try-on is the lack of datasets containing explicit size annotations or ill-fitting examples. Standard 2D benchmarks, such as ViTON [Han et al. 2018], ViTON-HD [Choi et al. 2021], DressCode [Morelli et al. 2022], StreetTryOn [Cui et al. 2023], and LAION-Garment [Guo et al. 2025], predominantly feature well-fitted garments, lacking the diverse fit conditions required for size-aware training. While some datasets, including SIZER [Tiwari et al. 2020] and SV-VTO [Yamashita et al. 2024], collect real-world samples for this purpose, they remain limited in scale and diversity (see Table 1).

Alternatively, 3D datasets [Bertiche et al. 2020; Liu et al. 2023; Tiwari et al. 2020; Zhu et al. 2020; Zou et al. 2023] offer 3D models of clothed humans. However, extracting accurate garment measurements from raw meshes is often infeasible. GarmentCode [Korosteleva and Sorkine-Hornung 2023] addresses this by introducing a domain-specific language for generating sewing patterns with explicit size parameters, enabling synthetic garment generation across varied garment and body sizes [Korosteleva et al. 2024]. However, for extreme ill-fitting garment draping cases, GarmentCode tends to produce significant and frequent draping errors. Furthermore, raw 3D synthetic datasets [Korosteleva et al. 2024; Li et al. 2025] generally suffer from their synthetic appearance, which leads to poor real-world generalization. Although Sewformer [Liu et al. 2023] attempts to enhance realism via texture synthesis and SDEdit refinement, the results remain clearly synthetic and lack fit diversity. To bridge these gaps, we adapt GarmentCode for ill-fit scenarios,

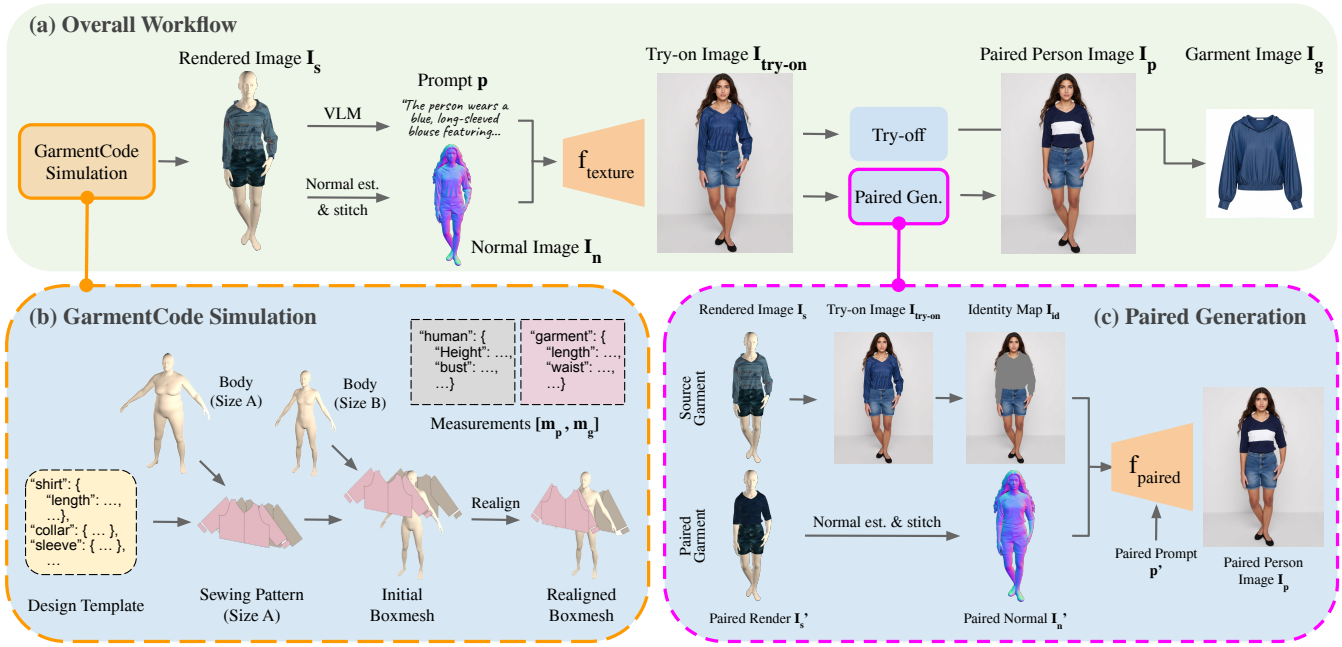


Fig. 2. **FIT data generation pipeline.** (a) Overall workflow: We start by simulating a 3D garment on a target body via GarmentCode to render a synthetic image $I_s$. We generate a text prompt $p$ (via VLM) and a composite normal map $I_n$ (stitching estimated normals with realistic head/feet details). These condition our re-texturing model $f_{\text{texture}}$ to produce the try-on image $I_{\text{try-on}}$. Finally, we use $f_{\text{paired}}$ to generate a paired person image $I_p$, and a VLM to synthesize a layflat garment $I_g$. (b) GarmentCode simulation: Given a sampled design template, we compute sewing patterns for a specific body size. Then, we cross-drape these patterns onto a different target body, using box-mesh realignment to prevent simulation failures, and extract ground-truth measurements. (c) Using source and target garments draped on the same body, we derive an identity map $I_{\text{id}}$ by masking the garment in $I_{\text{try-on}}$. Conditioned on $I_{\text{id}}$, a paired normal map $I_n'$, and a paired prompt $p'$, $f_{\text{paired}}$ generates the paired person image $I_p$.

as well as introduce a novel pipeline for transforming synthetic GarmentCode renderings into photorealistic images.

Table 1. Comparison of related datasets. We compare FIT to several related datasets. For scale, we report the number of training images.

| Dataset | Realism | Ill-Fit | Size Info | Triplet | Scale |
|---|---|---|---|---|---|
| SV-VTO | ✓ | ✓ | ✓ | ✓ | 1,524 |
| SIZER | ✓ | ✓ | ✓ | ✗ | 2,000 |
| DeepFashion3D | ✗ | ✗ | ✗ | ✗ | 2,078 |
| ViTON-HD | ✓ | ✗ | ✗ | ✗ | 11,647 |
| LAION-Garment | ✓ | ✗ | ✗ | ✓ | 60K |
| SewFactory | ✓ | ✗ | ✓ | ✗ | 1M |
| GCD | ✗ | ✗ | ✓ | ✗ | 115K |
| **Ours** | ✓ | ✓ | ✓ | ✓ | 1.13M |

## 2.2 Image-Based Virtual Try-On

Image-based virtual try-on methods are generally categorized into two paradigms: *mask-based*, which utilize explicit segmentation maps to localize generation, and *mask-free*, which synthesize results directly without segmentation priors.

*Mask-Based Methods.* These approaches formulate virtual try-on as a conditional inpainting task, where the target clothing region is masked and filled based on the garment image and human priors. Early warping-based works [Choi et al. 2021; Han et al. 2018] established a two-stage paradigm: warping the garment to the target body followed by refinement. Recent approaches have shifted toward single-stage diffusion-based architectures, achieving state-of-the-art photorealism [Chong et al. 2024; Cui et al. 2023; Kim et al. 2024; Xu et al. 2025; Zhu et al. 2024, 2023]. However, because these methods rely on inpainting within a fixed mask, they primarily focus on texture preservation and body alignment, largely neglecting the physical reality of garment sizing.

*Mask-Free Methods.* Another line of research [Du et al. 2023, 2025; Ge et al. 2021a,b; Guo et al. 2025; Issenhuth et al. 2020; Zhang et al. 2025] focus on mask-free architectures. Since real-world paired data is unavailable, these methods typically rely on generating "pseudo-triplets" via generative modeling to enable supervised training. A common strategy involves a "Teacher-Student" distillation framework, where a mask-based "teacher" model swaps garments on training images to generate synthetic ground-truth for a mask-free "student". Similarly, Any2AnyTryOn [Guo et al. 2025] leverages a pre-trained inpainting model to digitally replace garments in the try-on region. A fundamental bottleneck is that this training data is itself hallucinated, causing models to inherit the artifacts and geometric inconsistencies of the teacher. In contrast, our synthetic pipeline simulates actual draping dynamics on 3D bodies, yielding true ground-truth pairs with precise geometry and segmentation, effectively bypassing the error accumulation of 2D pseudo-triplet generation.

*Fit and Size Control.* While most VTO works ignore size, a few attempts have been made to incorporate fit information using geometric heuristics. For instance, [Chen et al. 2023] leverages clothing landmarks to transform garment size, while [Kuribayashi et al. 2023] uses body-to-clothing ratios to resize the conditioning segmentation maps. More recently, [Yamashita et al. 2024] introduced a diffusion model conditioned on coarse fit descriptors (e.g., "tight" or "loose"). However, by relying on imprecise intermediate values or coarse labels, these methods struggle to generalize to complex poses and lack precise control. In contrast, our fit-aware model avoids noisy geometric heuristics by conditioning on exact metric measurements.

## 3 Fit-Inclusive Try-on (FIT) Dataset

In this section, we describe the construction of the FIT dataset. We first report the dataset statistics in Section 3.1. We then detail our data generation pipeline, illustrated in Figure 2, which consists of the following steps: (1) procedurally generating garment assets with measurements $m_g$ and simulating their drape across diverse sizes of bodies with measurements $m_p$ via GarmentCode (Section 3.2); (2) transforming the synthetic renderings $I_s$ into photorealistic try-on images $I_{\text{try-on}}$ via a geometry-preserving re-texturing framework (Section 3.3); (3) leveraging identity conditioning to generate a paired person reference image $I_p$ featuring the same person wearing a different garment (Section 3.4); and (4) synthesizing the corresponding layflat garment image $I_g$ using an off-the-shelf VLM model [Google 2025a] (Section 3.5).

### 3.1 Dataset Statistics

Our dataset consists of 1,137,282 training and 1000 test samples, each consisting of $(I_{\text{try-on}}, I_{\text{p}}, I_g, m_p, m_g)$. Our data covers 168 distinct body shapes (82 men's, 86 women's) in sizes XS-3XL, 528 body poses, as well as 158,483 unique top and garment designs. Our dataset covers a diverse range of fits, from loose to tight fits. We provide a histogram of each person/garment size combination in the appendix. The test dataset is balanced to match the overall distribution over gender, body sizes, and person/garment size combinations. See Figure 1 and the appendix for examples of our dataset.

### 3.2 GarmentCode Simulation

GarmentCode [Korosteleva and Sorkine-Hornung 2023] is a parametric programming framework that enables the procedural generation and draping of 3D garment patterns, allowing for precise control over sizing and design details.

To generate try-on images with diverse fits, we implement a cross-draping strategy. We begin by sampling various garment templates and human body models with known measurements $m_p$. From a garment template, we generate sewing patterns fitted to multiple human bodies of varying sizes. We then simulate draping these sewing patterns onto a single target human model via GarmentCode's custom implementation of Warp [Macklin 2022], thereby creating realistic "tight" and "loose" fit scenarios. However, direct cross-draping initially fails because the 3D box-mesh specified by standard sewing patterns is aligned with its original target body, causing severe misalignments when applied to a new body. We address this by explicitly realigning the initial box-mesh panels to the target mesh position before simulation. Please refer to the appendix for details. Furthermore, GarmentCode's default implementation stitches top and bottom garments together into a unified mesh, preventing the appearance of "tucked-out" shirts. We modify this behavior to drape the top and bottom garments in two separate steps (typically simulating the bottom garment first) to ensure proper layering and realistic interactions between items. The draped 3d mesh is then reposed and rendered with different person poses to form a synthetic rendering image $I_s$.

Our procedural framework allows us to programmatically extract precise ground-truth garment measurements $m_g$ in centimeters directly from the 2D sewing pattern specifications. We focus on five critical metrics used in standard sizing: garment length (high point shoulder to hem), bust circumference (width), and sleeve length for tops; and waist and out-seam length for bottoms. We also derive four key body measurements directly from GarmentCode's parametric body model: height, bust, waist, and hips.

### 3.3 Synthetic-to-Photorealistic Retexturing

Our re-texturing pipeline is designed to transform the synthetic rendering $I_s$ into a photorealistic image while strictly preserving the geometry of both the garment and the subject. Due to the lack of paired synthetic-to-real training data, we utilize surface normal maps as a geometry-preserving bridge between domains. Specifically, we fine-tune a diffusion model, $f_{\text{texture}}$ (based on Flux.1-dev [Black Forest Labs 2024]), to synthesize photorealistic textures conditioned on an input normal map $I_n$ and a text prompt $p$. $f_{\text{texture}}$ is trained on real-world images with the following objective:

$$\hat{I}_{\text{try-on}} = f_{\text{texture}}(I_n, p) \tag{1}$$

where $I_n = N(I_{\text{try-on}})$ represents the normal map extracted from an off-the-shelf estimator $N$ [Khirodkar et al. 2024].

Despite utilizing normal maps, a significant domain gap persists between real-world and synthetic data. First, synthetic renderings $I_s$ lack anatomical details, featuring bald heads and bare feet. To address this, we employ a composite refinement strategy: we prompt Nano Banana Pro [Google 2025b] to inpaint realistic facial features, hair, and footwear onto $I_s$, estimate the normals of this enhanced image, and stitch the resulting head and feet regions onto the original synthetic normal map. This ensures realistic semantic cues while leaving the body and garment geometry untouched. Second, synthetic meshes lack intricate surface details, such as pockets, buttons, or seams. We observe that our model $f_{\text{texture}}$ successfully hallucinates these details when guided by appropriate text prompts. We further align the domains by augmenting the training data with random normal map blurring. This simulates the smoothness of synthetic normal maps and improves generation quality.

### 3.4 Paired Person Reference Image Generation

Our synthetic framework enables the generation of ground-truth paired data by exploiting procedural controllability. By fixing the subject's shape and pose while draping two distinct garments, we obtain pairs of synthetic renderings $(I_s, I'_s)$, normal maps $(I_n, I'_n)$, garment masks $(m_g, m'_g)$, and prompts $(p, p')$.

First, we generate the primary try-on image $I_{\text{try-on}} = f_{\text{texture}}(I_n, p)$ using the re-texturing pipeline described in Section 3.3. Next, to

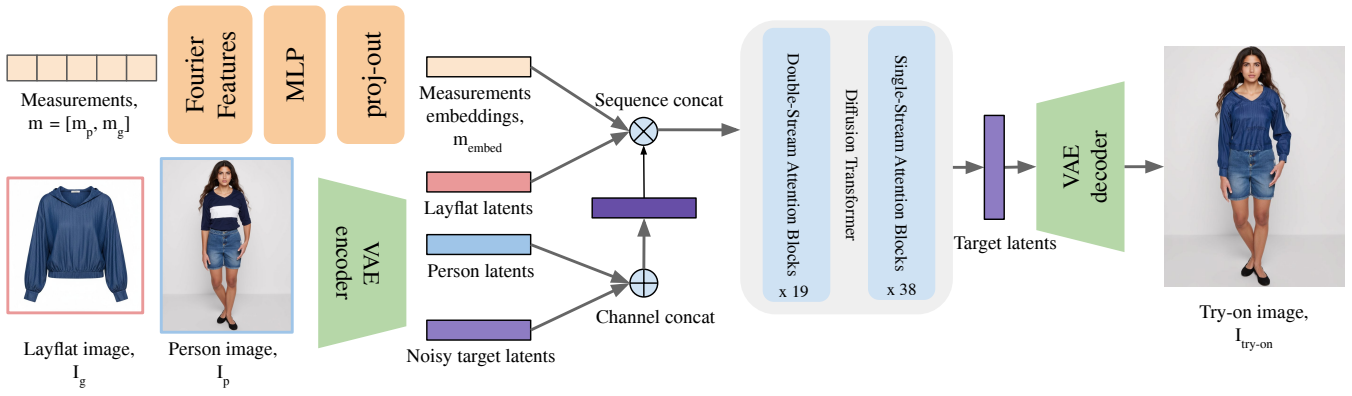


Fig. 3. Fit-VTO architecture. Our architecture is a flow-based diffusion model based on Flux.1-dev [Black Forest Labs 2024] and finetuned with LoRA [Hu et al. 2023]. FiT-VTO generates a try-on image $I_{\text{try-on}}$ given a layflat garment image $I_g$, paired person image $I_p$, and person-garment measurements $m = [m_p, m_g]$. First, image inputs $I_g$ and $I_p$ are encoded into latents separately through a pre-trained VAE encoder. We replace the text embeddings in Flux.1-dev with custom measurement embeddings $m_{\text{embed}}$ computed from $m$. Person latents are channel-concatenated with the noisy target latents, while layflat latents and $m_{\text{embed}}$ are sequence-wise concatenated with $z_t$. After processing through the diffusion transformer, clean latents are decoded by the VAE decoder.

synthesize the paired reference image $I_p$, we employ a conditional inpainting model $f_{\text{paired}}$:

$$I_p = f_{\text{paired}}(I_{\text{id}}, I'_n, p'), \tag{2}$$

where $I_{id}$ represents the identity map, defined as $I_{id} = I_{\text{try-on}} \odot (\neg m_g \cap \neg m'_g)$. This operation preserves the skin and background from the try-on image while masking out the regions occupied by both the source and paired garments. Essentially, $f_{\text{paired}}$ serves as a geometry-guided inpainter. To train $f_{\text{paired}}$, we utilize real human images, creates identity maps by estimating garment masks and applying random dilation to mimic the dual-garment masking seen at inference. Additionally, we limit our scope to upper-body try-on, hence we enforce identical bottom garment geometry across pairs during simulation. In practice, we train a unified model for $f_{\text{texture}}$ and $f_{\text{paired}}$ following Eq. 2, but randomly dropping out $I_{\text{id}}$.

### 3.5 Layflat Image Generation

Motivated by the impressive image synthesis capability of Nano Banana Pro [Google 2025b], we use it as an off-the-shelf virtual try-off model to generate a layflat garment image $I_g$ from $I_{\text{try-on}}$. Please refer to the appendix for the exact prompts used.

## 4 Fit-Aware Virtual Try-On

Given an image $I_p$ of person $p$, a garment image $I_g$ of target garment $g$, target garment measurements $m_g$, and person measurements $m_p$, our Fit-VTO model $f_{\text{vto}}$ synthesizes the predicted try-on result $\hat{I}_{\text{try-on}}$ of person $p$ wearing $g$ according to the measurements $m_p$ and $m_g$.

$$\hat{I}_{\text{try-on}} = f_{\text{vto}}(I_p, I_g, m_p, m_g) \tag{3}$$

### 4.1 Dataset Preparation

To increase the robustness of our model to diverse, real-world garments and poses, we crawled 330,559 online fashion images and their corresponding layflat garment images $I_g$, to augment our FIT training dataset. See the appendix for details. Since ground-truth measurements are not available for online images, we set the measurements to null values (−1). For FIT data samples, all measurements $m_p$, $m_g$ are normalized between 0 and 1.

### 4.2 Architecture

Our architecture (Figure 3) is a flow-matching diffusion model $x_\theta$ represented as:

$$\hat{v}_t = x_\theta(z_t, t, I_p, I_g, m_p, m_g) \tag{4}$$

where $z_t$ is the noisy ground-truth image $x_0$ at diffusion timestep $t$ and $\hat{v}_t$ is the predicted velocity. The model $x_\theta$ is trained to satisfy the consistency constraint where $\hat{v}_t$ approximates ground-truth velocity $v_t = x_0 - z_0$.

Our network $x_\theta$ is finetuned from the pre-trained Flux.1-dev text-to-image model [Black Forest Labs 2024]. FLUX.1-dev is a powerful, 12 billion parameter text-to-image generator that employs a rectified flow formulation and a Multi-modal Diffusion Transformer (MMDiT) backbone for efficient, high-fidelity image synthesis. We finetune only the lightweight LoRA parameters, keeping the majority of the original model weights frozen.

**Person and Garment Conditioning.** We condition the model on paired person image $I_p$ and garment image $I_g$. Since the $I_p$ is pixel-aligned to the noisy target image $z_t$, we concatenate latents from $I_p$ and $z_t$ channel-wise. Since $I_g$ latents need to be warped to $z_t$, we concatenate them along the sequence dimension after packing.

**Measurement Conditioning.** To condition on person and garment measurements, we remove the CLIP and T5 text conditionings for Flux.1-dev and instead condition with measurement embeddings from our custom measurement encoder $\mathcal{E}_m$. We first concatenate person measurements $m_p$ with garment measurements $m_g$ into a measurement vector $m = [m_p, m_g] \in R^7$. Then, we compute the Fourier Feature Embeddings for each measurement with 8 Fourier frequency bands, mapping $m \rightarrow m_{\text{embed}} \in R^{7\times16}$. These embeddings are further processed by an MLP and projected to the hidden dimension $R^{3072}$ of the MMDiT. Our model is conditioned on $m_{\text{embed}}$ with positional encodings for each measurement via cross-attention,

replacing the T5 text conditioning in the single-stream and double-stream blocks.

## 5 Experiments

We describe details of experiments in this section. We quantitatively and qualitatively evaluate the quality of our synthetic triplet data and demonstrate the effectiveness of our baseline fit-aware VTO model against state-of-the-art methods.

### 5.1 Implementation Details

For synthetic data generation, we initialize our re-texturing model from the pre-trained Flux.1-dev [Black Forest Labs 2024] checkpoint and only finetune with LoRA layers [Hu et al. 2023]. The model is trained on a custom dataset of 50k real person images (see appendix for details). We adopt Prodigy optimizer with learning rate 1.0 and weight decay factor 0.01. The training is done on 8 H200 GPUs with a total batch size of 64 and 5k training iterations.

Our baseline VTO model initialized from Flux.1-dev checkpoint. The measurement encoder is zero-initialized for stable early training. We fine-tune our model for 2M iterations on a mix of FIT training dataset and real-world images. The learning rate is $10^{-4}$ with 1000 warm-up steps and batch size is 64. All training is done on 64 TPU-v5's. At inference, we set guidance scale to 1.0 and number of inference steps is set to 50. We keep the same inference scheduler as the base Flux.1-dev release.

### 5.2 Evaluation Baselines, Datasets & Metrics

**Paired Image Generation Evaluation.** In this work, we propose a novel framework for generating pseudo-ground-truth paired-person images to enable mask-free VTO training. We benchmark against three baseline strategies: (1) **VLM-based**, which prompts Large Vision Language Models to swap garments while preserving context; (2) **VTO-based**, which utilizes off-the-shelf virtual try-on models for garment transfer; and (3) **Inpainting-based**, which replaces masked garment regions via generative inpainting. We implement these baselines using Nano Banana Pro, CatVTON [Chong et al. 2024], and FLUX-Controlnet-Inpainting [alimama-creative 2024].

To quantify how well the paired image $I_p$ preserves the identity of the original $I_{\text{try-on}}$ in non-garment regions (i.e., background, head, and limbs), we compute the Masked L1 Distance $\mathcal{L}_{\text{id}}$:

$$\mathcal{L}_{\text{id}} = \frac{1}{N} \sum \left| (I_p - I_{\text{try-on}}) \odot M \right|, \tag{5}$$

where $M = \mathbf{1} - (m_g \cup m'_g)$ represents the binary mask of the non-garment regions, and $N$ is the number of valid pixels in $M$. We randomly sampled 1000 cases from the dataset to compute this metric. The pixel values are between 0 and 255.

**Fit-Aware VTO Evaluation.** We compare our method qualitatively and quantitatively to Any2AnyTryon [Guo et al. 2025], Nano Banana Pro [Google 2025b], COTTON [Chen et al. 2023], IDM-VTON [Choi et al. 2024], and ablated versions of our method. We provide implementation details about related methods in the appendix. We evaluate on the VITON-HD test dataset [Choi et al. 2021] to measure general try-on accuracy and the FIT test dataset to evaluate fit-aware try-on accuracy. For VITON-HD, we generate paired-person images according to Section 3.4.

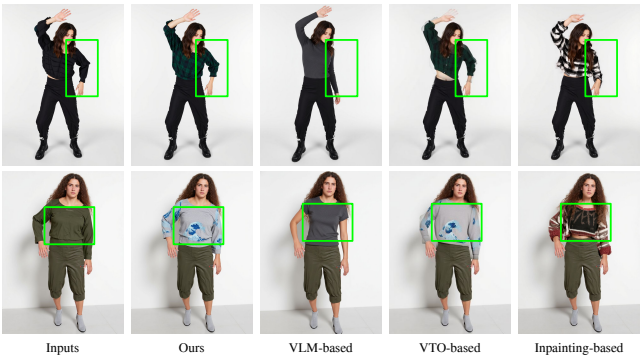


Fig. 4. Paired Image Generation Comparison. VLM methods struggle with pose and shape preservation, while VTO and inpainting baselines introduce artifacts. Our approach yields highly consistent paired data.

We compute common VTO metrics – SSIM [Ndajah et al. 2010], FID [Heusel et al. 2017], LPIPS [Zhang et al. 2018], KID [Binkowski et al. 2018] – to evaluate image similarity between ground-truth and synthesized try-on images. We also implement a custom metric (IoU), specifically designed for measuring size fidelity for the FIT dataset. IoU measures the Intersection-Over-Union of the garment mask in synthesized try-on image and the ground truth. We do not compute IoU for VITON-HD, as this dataset does not provide any size conditioning.

### 5.3 Paired Image Evaluation Results

We present a qualitative comparison of the generated paired-person images in Figure 4. Despite their impressive editing capabilities, VLM-based methods fail to guarantee identity preservation in non-garment regions (e.g., the left arm pose deviation in the top row). Furthermore, they often disregard the underlying body shape within the garment region (e.g., the inconsistent chest volume in the bottom row). Similarly, the VTO and Inpainting-based baselines introduce significant visual artifacts and struggle to maintain geometric consistency. In contrast, our approach achieves near-perfect identity and body shape preservation by explicitly conditioning on the identity map and ground-truth normals. Quantitative analysis confirms our visual findings: our method achieves an $\mathcal{L}_{\text{id}}$ of 1.61, significantly outperforming VLM-based (4.45), VTO-based (2.29), and Inpainting-based (3.91) baselines. These results demonstrate that our pipeline successfully generates highly consistent paired data essential for robust VTO training.

### 5.4 Fit-VTO Qualitative Results

We showcase qualitative results of Fit-VTO on the FIT test dataset in Figure 5. Our method synthesizes high-quality try-on images that maintain high fidelity to the person identity and garment appearance, while also synthesizing realistic garment fit with respect to the person and garment sizes. Fit-VTO handles diverse fit cases, including tight fit, perfect fit, and loose fit. We also show how the resulting garment fit responds to manual adjustments in individual

Table 2. Quantitative comparisons. We compare Fit-VTO to related methods and ablated versions of our method. $\text{Ours}_{\text{ft\_vitonhd}}$ refers to our method finetuned with VITON-HD training data. **Bolded** and underlined values indicate the best and second-best scores per column, respectively.

| | VITON-HD | | | | FIT Dataset | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | SSIM ↑ | FID ↓ | LPIPS ↓ | KID ↓ | SSIM ↑ | FID ↓ | LPIPS ↓ | KID ↓ | IOU ↑ |
| Any2AnyTryon [Guo et al. 2025] | 0.758 | 14.186 | 0.152 | 2.413 | 0.819 | 25.059 | 0.209 | 3.939 | 0.783 |
| Nano Banana Pro [Google 2025a] | 0.552 | 11.344 | 0.501 | 0.624 | 0.785 | 19.926 | 0.166 | 1.676 | 0.792 |
| COTTON [Chen et al. 2023] | 0.615 | 39.117 | 0.349 | 11.397 | 0.759 | 29.716 | 0.207 | 6.269 | 0.739 |
| IDM-VTON [Choi et al. 2024] | **0.849** | **9.115** | **0.077** | **0.471** | 0.739 | 31.229 | 0.246 | 6.819 | 0.789 |
| $\text{Ours}_{\text{no FIT}}$ | 0.817 | 11.499 | 0.103 | 0.639 | 0.852 | 16.427 | 0.095 | 0.849 | 0.844 |
| $\text{Ours}_{\text{text}}$ | 0.763 | 11.367 | 0.134 | 0.766 | 0.911 | 11.624 | 0.054 | 0.576 | 0.932 |
| $\text{Ours}_{\text{FIT only}}$ | 0.732 | 14.651 | 0.192 | 1.061 | 0.912 | 11.248 | 0.052 | 0.532 | 0.952 |
| **Ours** | 0.817 | 11.391 | 0.102 | 0.651 | **0.914** | **10.381** | **0.050** | **0.144** | **0.955** |
| $\textbf{Ours}_{\textbf{ft\_vitonhd}}$ | 0.833 | 9.320 | 0.087 | 0.670 | 0.846 | 17.041 | 0.096 | 1.724 | 0.908 |

garment measurements in Figure 7. See the appendix for additional Fit-VTO resizing results on real-world humans.

In Figure 6, we qualitatively compare our method to related works. Despite accurate texture warping, Any2AnyTryon [Guo et al. 2025], Nano Banana Pro [Google 2025b], COTTON [Chen et al. 2023], and IDM-VTON [Choi et al. 2024] fail to accurately portray accurate garment fit according to the person and garment sizes. COTTON also suffers from severe boundary artifacts due to errors in its pre-processing pipeline. In contrast to related methods, Fit-VTO accurately visualizes size and garment appearance.

### 5.5 Fit-VTO Quantitative Results

We compare quantitative metrics for related methods and Fit-VTO in Table 2. Fit-VTO excels in nearly all VTO metrics on both VITON-HD and FIT datasets compared to related methods. With additional finetuning, Fit-VTO even achieves comparable performance with IDM-VTON, a general try-on method, on the VITON-HD dataset. Plus, Fit-VTO achieves superior size-aware IoU score on the FIT dataset, even compared to size-conditioned COTTON. These results indicate that our method can simultaneously deliver high appearance fidelity, as well as incorporate size information for try-on.

### 5.6 Ablations

In our ablations, we evaluate the impact of our FIT dataset, measurements encoder, and real-world data supervision. As summarized in Table 2, we compare (1) training without FIT data and only online fashion images ($\text{Ours}_{\text{no FIT}}$), (2) replacing our measurement encoder with pre-trained T5 [Raffel et al. 2020] and CLIP [Radford et al. 2021] text encoders used in the original Flux.1-dev [Black Forest Labs 2024] model ($\text{Ours}_{\text{text}}$), and (3) training with FIT data only ($\text{Ours}_{\text{FIT only}}$). See Figure 6 for qualitative comparisons.

The FIT-only model performs well on FIT data, but degrades considerably on VITON-HD, as further evidenced in the bottom two rows of Figure 6. We attribute this to overfitting to the garments and poses FIT dataset, highlighting the importance of real-world training data for generalization. Conversely, the model trained without FIT data performs well on VITON-HD with respect to SSIM, FID, and LPIPS, but fails to model person-garment size relationships, as indicated by the significantly lower size-aware IoU. This demonstrates that real-world data with measurements predicted by VLM alone are insufficient for learning accurate fit. The text-only model performs moderately well on VITON-HD – likely because it better preserves the pretrained knowledge from Flux.1-dev – yet, this model fails to encode precise measurement information and exhibits a low IoU score. This indicates that pre-trained text encoders are not well-designed to represent structured numerical size inputs. Row 1 in Figure 6 corroborates these findings: $\text{Ours}_{\text{no FIT}}$ and $\text{Ours}_{\text{text only}}$ exhibit significant errors in representing ill-fitting garment size, while our full method accurately show accurate garment fit.

Our full model achieves the best balance across both benchmarks, performing on par with the strongest variants on each domain while delivering high size-aware IoU on FIT. These results confirm that combining FIT supervision, real-world data, and our measurement encoder yield a model that is both robust to real imagery and sensitive to garment–person size relationships.

## 6 Scope and Limitations

Our work serves as a proof-of-concept demonstrating that synthetic data generation, grounded in physics-based simulation, is a promising way to overcome the scarcity of size-annotated data in virtual try-on. However, as an initial exploration, our current scope is intentionally constrained. We focus exclusively on upper-body garments in standardized front-facing views (full-body or cropped) and casual poses, thereby avoiding the complicated collision dynamics. Additionally, the structural diversity of our dataset is bounded by the capabilities of the GarmentCode engine, limiting our study to simple structural designs rather than complex, multi-layered apparel. Despite these constraints, our results validate the core hypothesis: that synthetic, physics-informed supervision can teach generative models to respect precise metric sizing. We believe this synthetic-to-real paradigm establishes a foundation for future research to scale up to complex, in-the-wild scenarios.

We also identify two specific technical limitations. First, accurately representing the degree of tightness in the data is challenging. While the feeling of wearing a tight garment or a very tight garment may vastly differ, the simulated appearance is almost identical – fitted to the skin. As such, our dataset and VTO model do not represent varying degrees of tightness well (see left column of Figure 5). Furthermore, our Fit-VTO model is sensitive to correlations in measurements, limiting its ability to independently alter single measurements. For example, an increase in width frequently leads to a slight increase in length and sleeve length, as well.

## 7 Conclusions and Future Work

In this paper, we introduce **FIT**, the first large-scale dataset and benchmark for *fit-aware* virtual try-on (VTO) consisting of over 1.13M samples. We also present **Fit-VTO**, a novel fit-aware VTO model designed to leverage FIT's rich person-garment size annotations. Across extensive comparisons to related and ablated methods, Fit-VTO demonstrates a clear advantage in modeling accurate garment fit, according to person and garment measurements.

**Future Work:** Our immediate next steps involve expanding the dataset scope beyond tops to include lower-body and full-body garments (e.g., pants, dresses). Additionally, while our current dataset supports basic pose variation, scaling up the diversity of poses and camera viewpoints remains a key objective to ensure robust performance across complex, real-world inputs.

## References


alimama-creative. 2024. Flux controlnet inpainting. https://huggingface.co/alimama-creative/FLUX.1-dev-Controlnet-Inpainting-Beta.

Hugo Bertiche, Meysam Madadi, and Sergio Escalera. 2020. CLOTH3D: Clothed 3D Humans. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Mikolaj Binkowski, Danica J. Sutherland, Michal Arbel, and Arthur Gretton. 2018. Demystifying MMD GANs. *ArXiv* abs/1801.01401 (2018). https://api.semanticscholar.org/CorpusID:3531856

Black Forest Labs. 2024. FLUX.1 [dev]: A 12 Billion Parameter Rectified Flow Transformer. Hugging Face Model Repository.

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)*.

Zheng Chong, Xiao Dong, Haoxiang Li, Shiyue Zhang, Wenqing Zhang, Xujie Zhang, Hanqing Zhao, and Xiaodan Liang. 2024. CatVTON: Concatenation Is All You Need for Virtual Try-On with Diffusion Models. arXiv:2407.15886 [cs.CV] https://arxiv.org/abs/2407.15886

Aiyu Cui, Jay Mahajan, Viraj Shah, Preeti Gomathinayagam, Chang Liu, and Svetlana Lazebnik. 2023. Street TryOn: Learning In-the-Wild Virtual Try-On from Unpaired Person Images. *arXiv preprint arXiv:2311.16094* (2023).

Chenghu Du, Shuqing Liu, Shengwu Xiong, et al. 2023. Greatness in simplicity: Unified self-cycle consistency for parser-free virtual try-on. *Advances in Neural Information Processing Systems* 36 (2023), 20287–20298.

Chenghu Du, Shengwu Xiong, Junyin Wang, Yi Rong, and Shili Xiong. 2025. Mitigating Occlusions in Virtual Try-On via A Simple-Yet-Effective Mask-Free Framework. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

Chongjian Ge, Yibing Song, Yuying Ge, Han Yang, Wei Liu, and Ping Luo. 2021a. Disentangled cycle consistency for highly-realistic virtual try-on. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16928–16937.

Yuying Ge, Yibing Song, Ruimao Zhang, Chongjian Ge, Wei Liu, and Ping Luo. 2021b. Parser-free virtual try-on via distilling appearance flows. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8485–8493.

Yuying Ge, Ruimao Zhang, Xiaogang Wang, Xiaoou Tang, and Ping Luo. 2019. DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Retrieval of Clothing Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Google. 2025a. Gemini 2.5 Flash Image. 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)*.

Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S. Davis. 2018. VITON: An Image-based Virtual Try-On Network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Wang, Yu Chen, Lu Li, Xiangru Wang, Lu Wang, Yan Zhou, et al. 2023. LoRA: Low-Rank Adaptation of Large Language Models. *arXiv preprint arXiv:2308.03303* (2023).

Thibaut Issenhuth, Jérémie Mary, and Clément Calauzenes. 2020. Do not mask what you do not need to mask: a parser-free virtual try-on. In *European Conference on Computer Vision*. Springer, 619–635.

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.

Jeongho Kim, Guojung Gu, Minho Park, Sunghyun Park, and Jaegul Choo. 2024. Stableviton: Learning semantic correspondence with latent diffusion model for virtual try-on. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8176–8185.

Jeongho Kim, Hoiyeong Jin, Sunghyun Park, and Jaegul Choo. 2025. Promptdresser: Improving the quality and controllability of virtual try-on via generative textual prompt and prompt-aware mask. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 16026–16036.

Maria Korosteleva, Timur Levent Kesdogan, Fabian Kemper, Stephan Wenninger, Jasmin Koller, Yuhan Zhang, Mario Botsch, and Olga Sorkine-Hornung. 2024. GarmentCodeData: A Dataset of 3D Made-to-Measure Garments With Sewing Patterns. In *European Conference on Computer Vision (ECCV)*. https://arxiv.org/abs/2405.17609

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

Minoru Kuribayashi, Koki Nakai, and Nobuo Funabiki. 2023. Image-based virtual try-on system with clothing-size adjustment. *arXiv preprint arXiv:2302.14197* (2023).

Siran Li, Ruiyang Liu, Chen Liu, Zhendong Wang, Gaofeng He, Yong-Lu Li, Xiaogang Jin, and Huamin Wang. 2025. GarmageNet: A Multimodal Generative Framework for Sewing Pattern Design and Generic Garment Modeling. *ACM Trans. Graph.* 44, 6, Article 216 (Dec. 2025), 23 pages. doi:10.1145/3763271

Lijuan Liu, Xiangyu Xu, Zhijie Lin, Jiabin Liang, and Shuicheng Yan. 2023. Towards Garment Sewing Pattern Reconstruction from a Single Image. *ACM Trans. Graph.* 42, 6, Article 200 (Dec. 2023), 15 pages. doi:10.1145/3618319

Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. 2016. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Miles Macklin. 2022. Warp: A high-performance python framework for gpu simulation and graphics. In *NVIDIA GPU Technology Conference (GTC)*, Vol. 3.

Davide Morelli, Matteo Fincato, Marcella Cornia, Federico Landi, Fabio Cesari, and Rita Cucchiara. 2022. Dress Code: High-Resolution Multi-Category Virtual Try-On. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Peter Ndajah, Hisakazu Kikuchi, Masahiro Yukawa, Hidenori Watanabe, and Shogo Muramatsu. 2010. SSIM image quality metric for denoised images. In *Proceedings of the 3rd WSEAS International Conference on Visualization, Imaging and Simulation* (Faro, Portugal) *(VIS '10)*. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 53–57.

Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. 2020. TailorNet: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamila Mishkin, Jack Clark, Gretchen Krueger, and Dario Amodei. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21, 140 (2020), 1–67.

Garvita Tiwari, Bharat Lal Bhatnagar, Tony Tung, and Gerard Pons-Moll. 2020. SIZER: A Dataset and Model for Parsing 3D Clothing and Learning Size Sensitive 3D Clothing. In *European Conference on Computer Vision (ECCV)*. Springer.

Yuhao Xu, Tao Gu, Weifeng Chen, and Arlene Chen. 2025. Ootdiffusion: Outfitting fusion based latent diffusion for controllable virtual try-on. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 8996–9004.

Yohei Yamashita, Chihiro Nakatani, and Norimichi Ukita. 2024. Size-Variable Virtual Try-On with Physical Clothes Size. *arXiv preprint arXiv:2412.06201* (2024).

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. arXiv:arXiv:1801.03924

Xuanpu Zhang, Dan Song, Pengxin Zhan, Tianyu Chang, Jianhao Zeng, Qingguo Chen, Weihua Luo, and An-An Liu. 2025. Boow-vton: Boosting in-the-wild virtual try-on via mask-free pseudo data training. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 26399–26408.

Heming Zhu, Yu Cao, Hang Jin, Weikai Chen, Dong Du, Zhangye Wang, Shuguang Cui, and Xiaoguang Han. 2020. Deep Fashion3D: A Dataset and Benchmark for 3D Garment Reconstruction from Single Images. arXiv:arXiv:2003.12753

Luyang Zhu, Yingwei Li, Nan Liu, Hao Peng, Dawei Yang, and Ira Kemelmacher-Shlizerman. 2024. M&M VTO: Multi-Garment Virtual Try-On and Editing. *CoRR* abs/2406.04542 (2024). arXiv:2406.04542 doi:10.48550/ARXIV.2406.04542

Luyang Zhu, Dawei Yang, Tyler Zhu, Fitsum Reda, William Chan, Chitwan Saharia, Mohammad Norouzi, and Ira Kemelmacher-Shlizerman. 2023. TryOnDiffusion: A Tale of Two UNets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4606–4615.

Xingxing Zou, Xintong Han, and Waikeung Wong. 2023. CLOTH4D: A Dataset for Clothed Human Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.


Fig. 5. Qualitative results. Fit-VTO respects person and garment inputs, while also synthesizing realistic garment fit based on person and garment measurements (zoom in for details). For brevity, we approximate the full measurements with a size label (XS-3XL). See the appendix for our size categorization chart.


Fig. 6. Qualitative comparisons. Compared to related methods and ablated versions, our method best depicts accurate garment appearance and fit.


Fig. 7. Qualitative resizing results. Fit-VTO adapts garment fit according to garment measurements. We show results of independently shrinking and growing the length, width, and sleeve length with respect to the original value. Please zoom in for details.