SIMGARMENT: END-TO-END LEARNING OF GARMENT SEWING PATTERNS VIA DIFFERENTIABLE SIMULATION

Anonymous authors
Paper under double-blind review

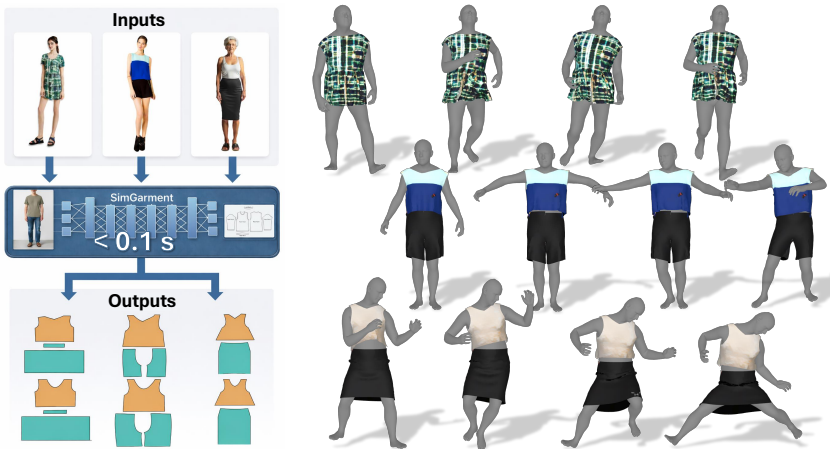


Figure 1: **SimGarment** predicts sewing patterns from a single full-body photo in under 0.1 s (left), producing structured sewing panels that can be simulated into physically plausible 3D garments across diverse subjects and garments (right).

ABSTRACT

We present SimGarment, the first method to integrate differentiable physics simulation directly into the training of a feed-forward garment reconstruction network. From a single-view image, SimGarment predicts 2D sewing panels and stitching using a Transformer model, assembles them into a garment mesh, and drapes it onto a parametric body via a differentiable simulator. This end-to-end differentiable pipeline allows gradients from 3D Chamfer and 2D silhouette losses to flow through the simulation stage, removing the requirement for ground-truth sewing panel annotations on all training data. This enables training on a mixture of synthetic data, real-world 3D scans, and in-the-wild RGB images, achieving strong generalization across garment categories unseen in synthetic training data. In zero-shot inference, SimGarment outperforms all feed-forward baselines by a wide margin and matches the quality of optimization-based methods, while being nearly five orders of magnitude faster, running in under 0.1 seconds per image.

Keywords: Sewing Pattern Reconstruction ; Differentiable Simulation

1 INTRODUCTION

Reconstructing garment sewing patterns from images is a fundamental problem with broad applications in fashion design, virtual try-on (Ning et al., 2024; Wang et al., 2024a; Choi et al., 2024), digital avatars, and content creation.

Recent research has advanced this problem through two main paradigms: optimization-based methods and feed-forward neural networks. Optimization-based approaches (Li et al., 2024b; 2025c; 2024a; Yu et al., 2024), such as Dress-1-to-3 (Li et al., 2025c), perform test-time optimization to fit sewing patterns to input images via physics simulation. While they produce reconstructions

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

that closely match the input image, the process is extremely slow, often taking hours per image, making them impractical for interactive or large-scale applications. Moreover, these methods lack strong structural priors and often yield unrealistic or distorted panel geometries that differ from artist-designed patterns. In contrast, feed-forward methods (Chen et al., 2024b; Nakayama et al., 2024; Liu et al., 2023) like SewFormer (Liu et al., 2023) enable fast inference by directly predicting sewing panels using Transformer architectures. However, they rely on ground-truth sewing panel annotations for supervision, restricting training to synthetic datasets with an inherent domain gap to real-world images. This limits their ability to leverage abundant real-world imagery and 3D scans, hindering generalization to in-the-wild garment styles, poses, and appearances.

In this work, we introduce SimGarment, the first method to integrate differentiable physics simulation directly into the training of a feed-forward garment reconstruction network. Given a single input image, a Transformer-based model predicts 2D garment panels and stitching, which are assembled into a 3D garment mesh and draped onto a parametric human body using a differentiable VBD simulator (Chen et al., 2024a). The simulation proceeds in two stages: (1) a sewing stage that pulls stitched panel boundaries together, and (2) a draping stage that simulates cloth dynamics under gravity and body contact. A key technical challenge arises from the non-differentiable triangulation step needed to convert predicted panel contours into a mesh. We address this by representing interior mesh vertices using harmonic coordinates (Li et al., 2024b; 2025c), continuous differentiable functions of the boundary vertices, allowing gradients to propagate seamlessly through the entire simulation pipeline. Gradients from observation-space losses flow back to the network via two supervision modes: (1) a bidirectional Chamfer loss for 3D scans, and (2) a silhouette loss from differentiable rendering for RGB images. This formulation enables SimGarment to train effectively on real-world data without ground-truth sewing panel annotations and achieve robust generalization.

Our main contributions are as follows:

- SimGarment is the first end-to-end framework integrating differentiable physics simulation into network training for fast feed-forward inference without test-time optimization.
- Our flexible training paradigm jointly leverages synthetic datasets with panel annotations and real-world 3D scans or RGB images without annotations, leading to strong generalization to in-the-wild conditions.
- SimGarment achieves reconstruction quality on par with optimization-based methods while being nearly five orders of magnitude faster (under 0.1 seconds vs. hours per image on a single A100 GPU), and significantly outperforms all feed-forward baselines on CloSe (Antić et al., 2024) and 4DDress (Wang et al., 2024b) benchmarks.

2 Related Work

2.1 Garment Reconstruction

Methods for recovering garment geometry from images either fuse garments with the body mesh (Xiu et al., 2022; 2023) or reconstruct independent garment representations. We focus on the latter, as separate representations support garment editing, simulation, and re-use.

Direct Mesh Reconstruction Several methods reconstruct garment meshes directly from images without an intermediate pattern representation. Weakly-supervised approaches such as ClothWild (Moon et al., 2022) and BCNet (Jiang et al., 2020) improve in-the-wild robustness but are constrained to their training garment categories. GarVerseLOD (Luo et al., 2024) recovers finer details via coarse-to-fine fitting, yet struggles with complex garment types and non-frontal poses. DeepFashion3D (Zhu et al., 2020) and AnchorUDF (Zhao et al., 2021) recover geometry from images but rely on dense annotations or produce coarse results without panel structure. Methods targeting simulation-ready outputs (Li et al., 2024b; Yu et al., 2024) incorporate differentiable physics but require artist-created templates or known 3D garment geometry as input. DrapeNet (De Luigi et al., 2023) learns a generative UDF-based garment model and trains a draping network purely via physics-based self-supervision, demonstrating that differentiable simulation can substitute for ground-truth draped data, but UDF representations lack panel structure and cannot directly support garment editing or manufacturing.

Sewing Pattern-Based Reconstruction Sewing patterns (Korosteleva & Sorkine-Hornung, 2023; He et al., 2024) offer an interpretable, editable representation for real-world garment manufacturing and physics simulation. Feed-forward methods learn to predict patterns directly from im-

ages. SewFormer (Liu et al., 2023) uses a hierarchical Transformer to jointly predict panel geometries and stitching. PanelFormer (Chen et al., 2024b) follows a similar architecture. NeuralTailor (Korosteleva & Lee, 2022) applies point-level attention to regress panel shapes and stitching from point clouds, but is constrained to point cloud inputs rather than RGB images. Recent works SewingLDM (Liu et al., 2024), AIpparel (Nakayama et al., 2024), ChatGarment (Bian et al., 2025), and Design2GarmentCode (Zhou et al., 2025) leverage multimodal foundation models for pattern synthesis. All of these methods share a fundamental limitation: they require ground-truth sewing pattern annotations for supervision, restricting training to synthetic datasets (Liu et al., 2023; Korosteleva & Sorkine-Hornung, 2023) and limiting generalization to real-world images with diverse garment styles.

An alternative is to fit a pattern representation to image or geometry observations at test time. Earlier works (Wang et al., 2018; Yang et al., 2018) jointly optimizes garment patterns and body pose in a shared latent space, but relies heavily on template datasets and generalizes poorly out-of-distribution. GarmentRecovery (Li et al., 2024a) and Li et al. (Li et al., 2025b) both leverage the Implicit Sewing Patterns (ISP) representation (Ren et al., 2023): GarmentRecovery fits an image-conditioned deformation prior to observations, while Li et al. use a diffusion model to map image pixels directly to ISP UV coordinates. Dress-1-to-3 (Li et al., 2025c) takes the predictions of SewFormer (Liu et al., 2023) as initialization and iteratively optimizes sewing patterns and material parameters via physics simulation to match the input image; while this yields physically plausible reconstructions, it requires approximately 2 hours per image, is prone to local minima, and cannot significantly deviate from the initial predictions.

Unlike traditional feed-forward approaches, SimGarment alleviates the need for ground-truth pattern annotations, facilitating training on diverse, non-synthetic datasets. By incorporating a differentiable physics simulation during training, it bridges the gap between fitting-based methods and neural networks, allowing for direct supervision from real-world observations while retaining an efficient feed-forward pipeline at inference.

2.2 Differentiable Simulation

Differentiable physics has been applied broadly to rigid bodies (Freeman et al., 2021), articulated systems (Xu et al., 2021), and deformable solids (Hu et al., 2019b; Du et al., 2021; Jatavallabhula et al., 2021), with cloth simulation being most relevant to our work (Li et al., 2022; Stuyck & Chen, 2023; Santesteban et al., 2022; Grigorev et al., 2023; Li et al., 2024b). A central challenge for differentiable cloth is handling frictional contact. Bartle et al. (Bartle et al., 2016) optimize garment patterns via fixed-point iteration without gradients. Liang et al. (Liang et al., 2019) pioneer fully differentiable cloth simulation with friction and self-collision via a quadratic program. DiffCloth (Li et al., 2022) and DiffPD (Du et al., 2021) use adjoint-based Projective Dynamics (Bouaziz et al., 2014); DiffXPBD (Stuyck & Chen, 2023) and DiffAvatar (Li et al., 2024b) extend Position-Based Dynamics (Müller et al., 2007; Macklin et al., 2016); and (Huang et al., 2024) develops an adjoint formulation for Incremental Potential Contact (Li et al., 2020; 2021).

For derivative computation, adjoint methods (Naumann, 2011; Margossian, 2019) are memory-efficient but assume full solver convergence, an assumption that often fails in practice with fixed iteration budgets. Direct differentiation via AD frameworks such as NVIDIA Warp (Macklin, 2022), DiffTaichi (Hu et al., 2019a), and related tools (Herholz et al., 2024) remains accurate under a fixed number of iterations, at the cost of storing intermediate simulation states. We adopt direct differentiation through a VBD solver (Chen et al., 2024a), which is efficient and numerically stable under limited iterations, making it well-suited for integration into a neural network training loop.

3 Method Overview

SimGarment trains a feed-forward sewing-pattern predictor through differentiable simulation (Fig. 2). Given an image, we recover an SMPL body, predict panels and stitching, mesh the panels, and simulate sewing and draping onto the body. Section 4 describes the differentiable simulator and contour-to-mesh mapping that connect the simulated garment to the network outputs.

Training combines three supervision sources (Section 5): synthetic panel and stitch annotations directly supervise network outputs; real 3D scans supervise simulated garments through Chamfer distance; and RGB images supervise rendered silhouettes through a mask loss. The real-data losses back-propagate through simulation, and rendering for image supervision, without requiring ground-truth sewing patterns.

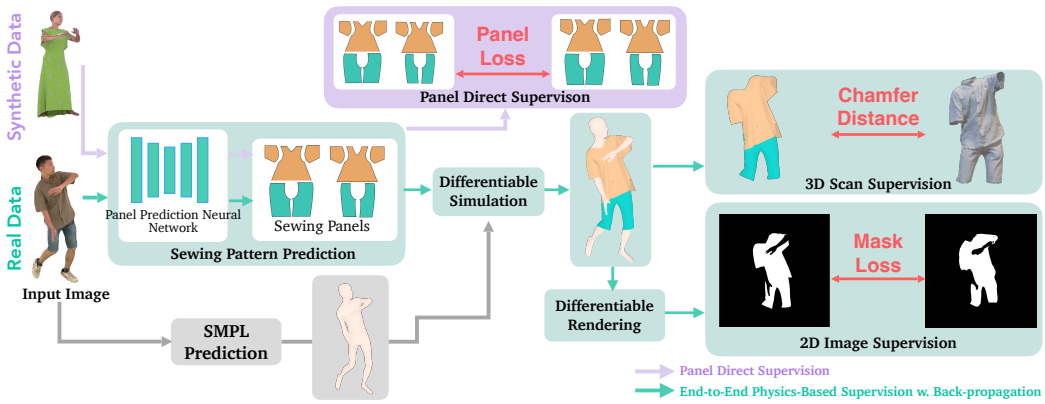


Figure 2: **SimGarment Training Pipeline.** Given a single input image, we predict sewing panels and an SMPL body, then drape the garments via differentiable simulation. When 3D scan ground truth is available, we apply Chamfer-distance supervision; otherwise, we differentiably render the simulated garment and supervise with a 2D silhouette mask loss. Gradients back-propagate through simulation and rendering to train the panel prediction network.

4 Differentiable Training Pipeline

This section presents the differentiable training pipeline used by SimGarment. We introduce a fast VBD-based differentiable garment-body simulator (Section 4.1) and a differentiable panel meshing formulation that enables gradients to pass through mesh triangulation step (Section 4.2).

4.1 Differentiable Garment-Body Simulator

A central component of SimGarment is a *training-time* garment-body simulator that is both fast and differentiable. Our goal is not only to produce physically plausible drapes, but also to provide stable gradients under a fixed, small iteration budget so that simulation can be embedded in the learning loop.

Implicit Cloth Dynamics. We formulate each time step using the standard variational form of implicit Euler:

$$x = \underset{x}{\operatorname{argmin}} \frac{1}{2h^2}\|x^- - y\|_M^2 + E_e(x) + E_c(x)\ , \tag{1}$$

where x^- is the previous state, y is the inertial prediction, and $E_e(x)$ and $E_c(x)$ denote elastic and collision energies. We use StVK-based shell elasticity with hinge-angle bending, and model garment-body contact with a quadratic energy defined on the body's signed distance field (SDF).

Fixed-iteration VBD solver for Stable Gradient. To make simulation practical during training, we solve Eq. 1 using a differentiable Vertex Block Descent (VBD) solver (Chen et al., 2024a). VBD is particularly well-suited to our setting because it remains stable and accurate even when run for a small, fixed number of iterations, which yields predictable runtime and avoids expensive full convergence. This property is critical for simulation-in-the-loop training at scale.

Differentiable VBD in Warp. We implement the solver in NVIDIA Warp (Macklin, 2022) and compute gradients by *direct differentiation* through unrolled VBD iterations using Warp's automatic differentiation engine. We prefer direct differentiation to an adjoint method because our forward solve does not enforce full convergence, making adjoint gradients inaccurate under early termination. Although direct differentiation stores intermediate states and uses more memory, the overhead is manageable at our mesh resolutions (a few thousand vertices) and yields robust training gradients. Our differentiable simulation is therefore **several orders of magnitude faster** than prior optimization-based pipelines (Li et al., 2025c), while remaining sufficiently accurate for training.

4.2 Back-propagation Through Differentiable Physics

Our differentiable simulator (Section 4.1) provides gradients of losses defined on the simulated drape with respect to the simulator inputs (panel meshes, material parameters, and stitch constraints). We now connect these gradients to the panel prediction network (Fig. 3) so that supervision from images and scans can directly update the predicted panel boundaries and stitching.

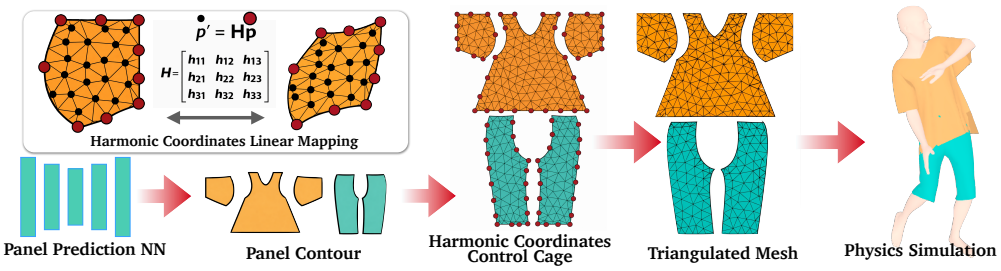


Figure 3: **Back-propagation through panel meshing.** Harmonic coordinates express interior vertices (black) as $p' = Hp$, a linear function of contour vertices (dark red). This mapping links the triangulated mesh to predicted contours so simulation-loss gradients update the panel predictor.

Contour-to-mesh Mapping. Given an input image, our network predicts (i) a set of 2D sewing panels and (ii) stitch correspondences between panel boundary segments. Each panel is represented as a closed planar curve composed of cubic Bézier segments, parameterized by endpoints and control points. For physics-based sewing and draping, we require a triangle mesh for each panel. However, standard contour triangulation is a discrete operation and therefore not directly differentiable. To enable end-to-end training, we establish a differentiable mapping from contour vertices to the vertices of the triangulated mesh, allowing gradients from the simulated deformation to propagate back to the predicted boundary (Li et al., 2024b; 2025c).

Harmonic-coordinate Parameterization. To make the contour-to-mesh step differentiable, we parameterize the triangulated panel mesh by its boundary. We sample B boundary vertices from the Bézier contour as $p \in \mathbb{R}^{2B}$, and denote the I interior mesh vertices from triangulation by $p' \in \mathbb{R}^{2I}$ (Fig. 3 inset). Using harmonic coordinates, we compute $p' = Hp$, where $H \in \mathbb{R}^{2I \times 2B}$ is a precomputed harmonic mapping matrix. This linear map yields a smooth differentiable dependence of triangulated vertex positions on the predicted contour. We assemble the mesh vertices as $V(p) = [p; p']$ and pass the resulting triangulated mesh to the differentiable simulator.

Gradient Flow. Let $\mathcal{L}$ denote a simulation loss (e.g., Chamfer distance or silhouette loss). The simulator provides $\partial\mathcal{L}/\partial V$, and we back-propagate to the panel contour via the harmonic map:

$$\frac{\partial \mathcal{L}}{\partial p} = \left(\frac{\partial p'}{\partial p}\right)^{\top} \frac{\partial \mathcal{L}}{\partial p'} = H^{\top} \frac{\partial \mathcal{L}}{\partial p'}.$$

Since p is a differentiable function of the network outputs (Bézier control points), these gradients propagate to the panel prediction network parameters, enabling simulation-in-the-loop training without ground-truth sewing patterns.

5 Learning with Heterogeneous Supervision

5.1 Panel Prediction Network

SimGarment builds on the panel prediction network from SewFormer (Liu et al., 2023). Given an input image, the model predicts a set of 2D sewing panels along with their stitching and body parameters. Panel contours are represented as boundary vertices in UV space, while a stitching matrix encodes which panel edges should be joined. These predictions serve as inputs to our differentiable simulation pipeline, where panels are converted into meshes, stitched along corresponding seams, and draped onto the body. We initialize SimGarment from a pre-trained SewFormer checkpoint and fine-tune end-to-end using the differentiable physics-based objectives below.

5.2 Physics-Based Garment Draping Simulation

We use the differentiable VBD simulator (Section 4.1) in two stages. Starting from panels placed around a T-posed SMPL body, **sewing** joins paired stitch vertices through spring constraints with gravity disabled. During **draping**, the body transitions from T-pose to the target pose while the garment settles under gravity and body contact. Initialization, component alignment, and the full procedure are provided in the supplementary section "Garment Initialization and Draping".

5.3 TRAINING OBJECTIVES

SimGarment uses different supervision signals depending on the data source. For synthetic data, we follow standard pattern-supervised losses from SewFormer (Liu et al., 2023). For real data, 3D scans and RGB images, we introduce novel physics-based losses that back-propagate gradients through the differentiable simulator to the network, requiring no ground-truth pattern annotations. Two regularizers improve physical plausibility and stitching quality across data sources.

Synthetic data. Although synthetic data exhibits a domain gap from real images, it provides ground-truth panel and stitch annotations that real-world data cannot supply. This direct supervision is essential for grounding the network in plausible panel geometries and stitch configurations, without which the physics-based losses alone may not sufficiently constrain the prediction space. Following SewFormer (Liu et al., 2023), we compute a panel prediction loss, a stitching prediction loss, and an SMPL joint loss on synthetic image–panel pairs.

Objectives for 3D scans. For 3D scan data, we back-propagate a geometry loss through the differentiable simulator to supervise the network without pattern annotations. Concretely, we simulate the predicted panels into a 3D garment and compute a bidirectional Chamfer distance against the ground-truth garment extracted from the scan:

$$\mathcal{L}_{\text{chamfer}}(G_{\text{gt}}, G_{\text{pred}}) = \frac{1}{|G_{\text{gt}}|} \sum_{x \in G_{\text{gt}}} \min_{y \in G_{\text{pred}}} \|x - y\|_2^2 + \frac{1}{|G_{\text{pred}}|} \sum_{y \in G_{\text{pred}}} \min_{x \in G_{\text{gt}}} \|y - x\|_2^2,$$

where G_{gt} denotes points sampled from the ground-truth garment surface and G_{pred} denotes the vertices of the simulated garment. Gradients of $\mathcal{L}_{\text{chamfer}}$ flow back through the simulator to update the pattern prediction network directly from 3D geometry observations.

Objectives for 2D images. For RGB images, we exploit our differentiable simulation and rendering pipeline to supervise the network from 2D observations alone, without requiring 3D scans or panel annotations. Given the simulated garment mesh $G_{\text{pred}} = (V, F)$, we recover per-view camera parameters and transform garment vertices into each camera k's coordinate system as $V_k = V R_k^\top + t_k$, then apply a differentiable renderer $\mathcal{R}$ (Ravi et al., 2020) to obtain a soft silhouette mask $\hat{S}_k = \mathcal{R}(V_k, F; K_k) \in [0,1]^{H \times W}$. The mask loss is the average MSE between the rendered and ground-truth segmentation masks S_k over all valid views $\mathcal{V}$: $\mathcal{L}_{\text{mask}} = |\mathcal{V}|^{-1} \sum_{k \in \mathcal{V}} \|\hat{S}_k - S_k\|_2^2$. Gradients flow end-to-end through both the differentiable renderer and simulator, enabling the network to learn from real images with only segmentation masks as supervision.

Regularizations. To improve physical plausibility and training stability, we apply two auxiliary regularizers to both the 2D-image-supervised and 3D-scan-supervised objectives. *Stretch regularization* $\mathcal{L}_{\text{area}}$ penalizes relative changes in triangle area beyond a tolerance, discouraging excessive expansion or compression during simulation. *Stitching regularization* $\mathcal{L}_{\text{stitch-ce}}$ uses cross-entropy to distill row-wise edge-similarity distributions from a frozen, pre-trained SewFormer teacher, providing stitching guidance when labels are unavailable.

Stretch regularization. We penalize excessive in-plane stretching after simulation. Let $\mathcal{T}$ denote the set of garment mesh triangles, and let A_t^0 and A_t^1 be the triangle areas at the initial (rest) state and final simulated frame, respectively. We define the relative area change r_t and stretch loss $\mathcal{L}_{\text{area}}$ with tolerance δ (e.g., $\delta = 0.1$):

$$r_t = \frac{|A_t^1 - A_t^0|}{A_t^0}, \qquad \mathcal{L}_{\text{area}} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} [\max(0, r_t - \delta)]^2$$

This loss is fully differentiable w.r.t. the simulated vertices and discourages unrealistic triangle-wise area expansion or compression.

Stitching regularization. Without ground-truth stitching supervision available from real datasets, the model tends to produce implausible panels or fail to stitch them correctly. We therefore distill stitching structure prior from a frozen, pre-trained SewFormer teacher. Let $M \in \mathbb{R}^{N \times N}$ be the edge similarity matrix predicted by the teacher and $\tilde{M} \in \mathbb{R}^{N \times N}$ be that of our model (student), with entries as cosine similarities in $[-1, 1]$. We convert teacher similarities into a probability distribution via row-wise softmax with temperature τ and treat $\tilde{M}$ as logits to obtain row-wise log-probabilities,

Table 1: Chamfer distance (mean $\pm$ std, $\downarrow$) on CloSe, 4DDress, and the combined benchmark. Methods are grouped by inference paradigm, with indicative per-image runtimes.

METHOD	CATEGORY	RUNTIME	PERFORMANCE		
			OVERALL	CLOSE	4DDRESS
GarmentRecovery	Optimization	$\sim$2 hours	2.216 ± 0.635	2.058 ± 0.412	2.659 ± 0.944
Dress1-to-3			1.592 ± 0.324	1.576 ± 0.280	$\mathbf{1.675} \pm 1.059$
BCNet	Feed-Forward	<0.1 sec	2.745 ± 1.349	2.244 ± 0.552	5.448 ± 1.668
ClothWild			2.213 ± 0.740	2.131 ± 0.598	2.658 ± 1.005
SewFormer			2.233 ± 0.605	2.214 ± 0.604	2.335 ± 1.099
Ours	Feed-Forward	<0.1 sec	$\mathbf{1.589} \pm 0.350$	$\mathbf{1.567} \pm 0.342$	1.706 ± 0.396

then minimize the row-wise cross-entropy between teacher and student, averaged over all rows and batch elements:

$$P = \mathrm{softmax}\left(\frac{M}{\tau}\right), \qquad \log Q \quad = \log \mathrm{softmax}\left(\frac{\tilde{M}}{\tau}\right).$$

$$\mathcal{L}_{\text{stitch-ce}} = -\frac{1}{B}\sum_{b=1}^{B}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} P_{ij}^{(b)} \log Q_{ij}^{(b)}.$$

This loss encourages the model to reproduce the stitching similarity structure of the pre-trained SewFormer, effectively regularizing stitching predictions on unannotated real data.

6 EXPERIMENTS

6.1 TRAINING DATA AND SUPERVISION

Our method is trained on three complementary data sources: synthetic data, 3D scans, and in-the-wild images. For **synthetic data**, we use the SewFactory dataset (Liu et al., 2023), which provides ground-truth panel and stitch annotations. Although synthetic data introduces a domain gap relative to real-world imagery, it offers strong structural priors that help the model learn plausible panel shapes and stitching patterns. For **3D scans**, we leverage CloSe (Antić et al., 2024) and 4DDress (Wang et al., 2024b) as sources of 3D garment meshes. The body pose and shape parameters in these datasets are often inaccurate, occasionally causing mesh interpenetration between the parametric body and clothing. We therefore re-estimate SMPL parameters using ETCH (Li et al., 2025a), which is specifically designed for clothed 3D scans. We curate **in-the-wild human photos** from DeepFashion (Liu et al., 2016) and generate corresponding multi-view samples following (Li et al., 2025c). Body pose and shape are estimated using GVHMR (Shen et al., 2024), and garment regions are segmented using SegFormer (Xie et al., 2021).

6.2 BASELINES

We compare against two optimization-based approaches (GarmentRecovery (Li et al., 2024a) and Dress-1-to-3 (Li et al., 2025c)) and three feed-forward methods (BCNet (Jiang et al., 2020), ClothWild (Moon et al., 2022), and SewFormer (Liu et al., 2023)). Optimization-based methods perform instance-specific fitting during inference, whereas all feed-forward methods, including ours, operate in a zero-shot manner. Since the official implementation of Dress-1-to-3 is not publicly available, we use garment meshes provided by the authors and align our reconstructed meshes to them before computing the Chamfer distance. See the supplement for implementation details.

6.3 QUALITATIVE RESULTS

Figure 4 compares garment reconstructions. BCNet and ClothWild exhibit category mismatches (red boxes) or inaccurate garment lengths (blue boxes), while SewFormer produces geometric artifacts (purple boxes). SimGarment better recovers the illustrated garment shapes and achieves visual quality comparable to Dress-1-to-3 without per-instance garment optimization.

Figure 5 compares sewing panels. In the first row, SewFormer predicts pants for a dress; this error persists in Dress-1-to-3, which optimizes panel boundaries while retaining SewFormer's initial

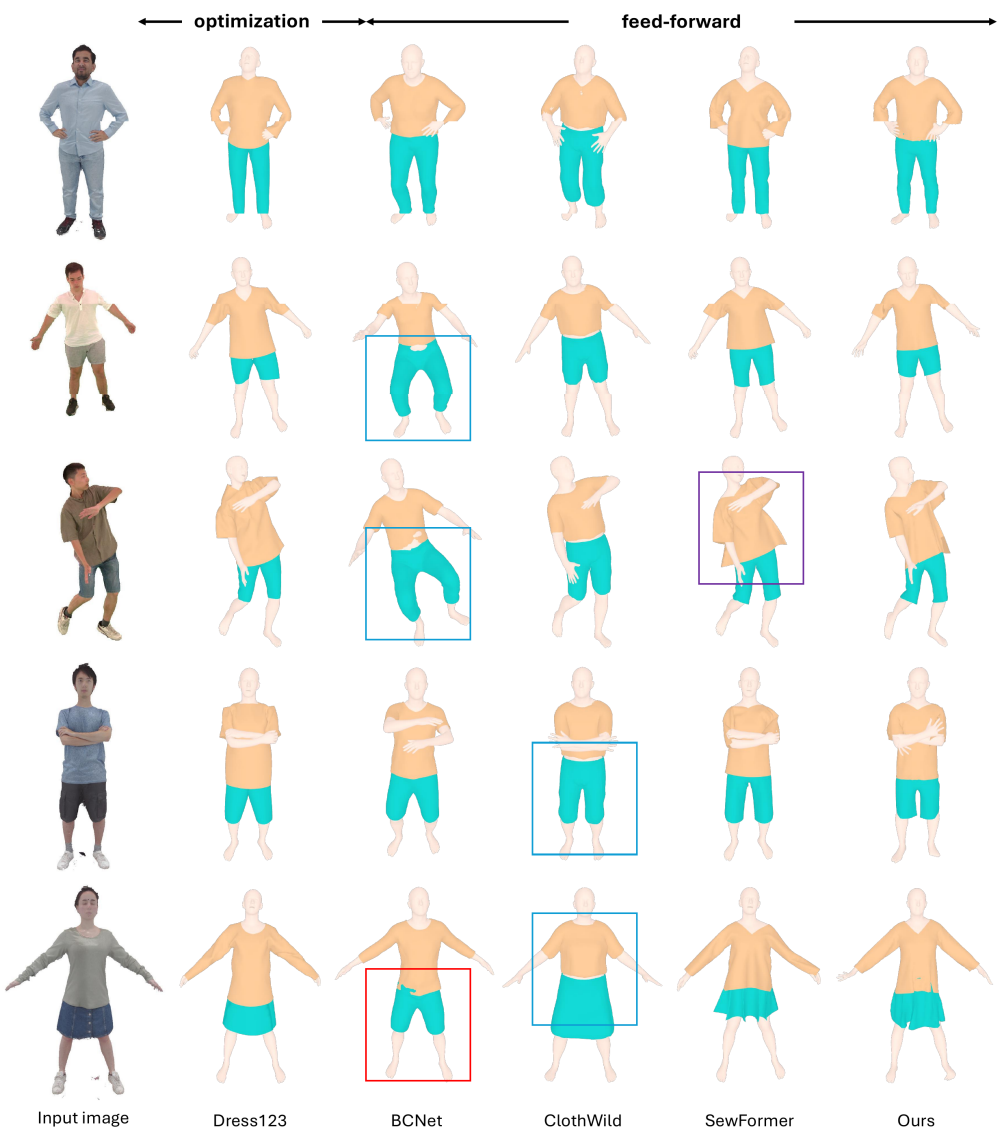


Figure 4: **Comparison with baseline methods.** Red, blue, and purple boxes highlight category, garment-length, and geometric errors, respectively. BCNet (Jiang et al., 2020), ClothWild (Moon et al., 2022), SewFormer (Liu et al., 2023), and SimGarment are feed-forward methods; Dress-1-to-3 (Li et al., 2025c) uses per-instance optimization.

topology. SimGarment recovers the dress structure and avoids the illustrated panel distortions (blue box) and sleeve/collar errors (purple boxes). The textured meshes in the second column are generated from the predicted garments and input images using Hunyuan3D (Team, 2025b;a; 2024).

6.4 Quantitative Evaluation

Following Dress-1-to-3, Table 1 reports Chamfer distance on CloSe (Antić et al., 2024) and 4DDress (Wang et al., 2024b). SimGarment achieves the lowest overall error among feed-forward methods (1.589, versus 2.213 for the strongest feed-forward baseline) and comparable overall error to Dress-1-to-3 (1.592). Relative to Dress-1-to-3, our error is slightly lower on CloSe and slightly higher on 4DDress. Unlike optimization-based approaches that fit garment parameters to each test image, SimGarment predicts sewing patterns in a single forward pass.

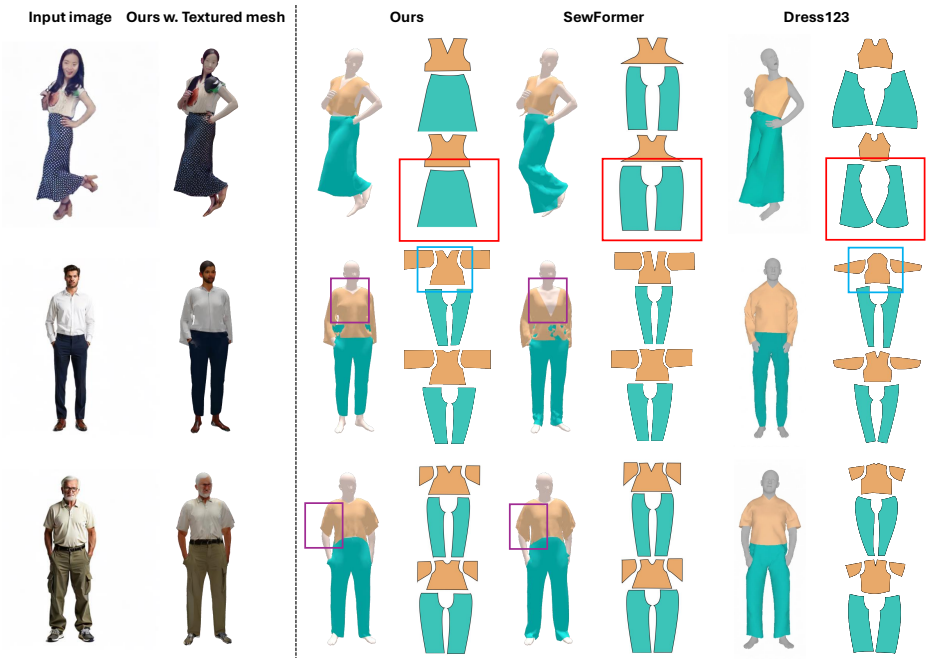


Figure 5: **Panel and texture reconstruction comparison.** SimGarment produces accurate textured meshes and plausible sewing panels, while SewFormer (Liu et al., 2023) and Dress-1-to-3 (Li et al., 2025c) exhibit category errors (red boxes), distorted panels (blue boxes), and inaccurate sleeve/collar geometry (purple boxes).

6.5 Ablation Studies

We ablate each regularization term from the full objective and evaluate on the combined benchmark. As shown in Table 2, both regularizers improve reconstruction accuracy. Removing stitching regularization substantially degrades performance, highlighting the importance of teacher-distilled stitch similarities when stitch labels are unavailable. Removing stretch regularization also worsens accuracy, confirming its role in discouraging unrealistic in-plane deformation and stabilizing simulation-driven gradients. Data-source ablations are provided in the supplementary section "Data Sources".

Table 2: Regularization ablation. Chamfer distance (mean $\pm$ std, $\downarrow$).

Model	CD $\downarrow$
Ours	$\mathbf{1.589 \pm 0.350}$
No Stitch	2.035 ± 1.341
No Area	2.076 ± 1.537

7 Conclusion

We presented SimGarment, a feed-forward framework for reconstructing simulation-ready sewing patterns from a single image. Differentiable simulation enables joint training on synthetic pattern annotations, real 3D scans, and RGB images without real-data pattern labels. The resulting model outperforms synthetic-only feed-forward baselines and achieves reconstruction quality comparable to test-time optimization, while predicting sewing patterns in under 0.1 seconds.

Limitations. SimGarment has several limitations. Cloth material parameters (stiffness, bending resistance) are fixed globally rather than predicted per garment, which limits fidelity for materials with highly disparate physical properties, such as stiff denim versus flowing chiffon. Additionally, garments with highly unconventional constructions (*e.g.*, asymmetric cuts, very high panel counts, multiple layers) may not be represented in current training data; extending to such cases through data augmentation or topology-flexible panel predictors is a natural direction for future work.

REFERENCES

Dimitrije Antić, Garvita Tiwari, Batuhan Ozcomlekci, Riccardo Marin, and Gerard Pons-Moll. Close: A 3d clothing segmentation dataset and model. In *2024 International Conference on 3D Vision (3DV)*, pp. 591–601. IEEE, 2024.

Aric Bartle, Alla Sheffer, Vladimir G Kim, Danny M Kaufman, Nicholas Vining, and Floraine Berthouzoz. Physics-driven pattern adjustment for direct 3d garment editing. *ACM Trans. Graph.*, 35(4):50–1, 2016.

Siyuan Bian, Chenghao Xu, Yuliang Xiu, Artur Grigorev, Zhen Liu, Cewu Lu, Michael J Black, and Yao Feng. Chatgarment: Garment estimation, generation and editing via large language models. *CVPR*, 2025.

Sofien Bouaziz, Sebastian Martin, Tiantian Liu, Ladislav Kavan, and Mark Pauly. Projective dynamics: Fusing constraint projections for fast simulation. *ACM transactions on graphics (TOG)*, 33(4):1–11, 2014.

Anka He Chen, Ziheng Liu, Yang Yin, and Cem Yuksel. Vertex block descent. *ACM Transactions on Graphics (TOG)*, (116), 2024a.

Cheng-Hsiu Chen, Jheng-Wei Su, Min-Chun Hu, Chih-Yuan Yao, and Hung-Kuo Chu. Panelformer: Sewing pattern reconstruction from 2d garment images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 454–463, 2024b.

Yisol Choi, Sangkyung Kwak, Kyungmin Lee, Hyungwon Choi, and Jinwoo Shin. Improving diffusion models for virtual try-on. *arXiv preprint arXiv:2403.05139*, 2024.

Luca De Luigi, Ren Li, Benoit Guillard, Mathieu Salzmann, and Pascal Fua. DrapeNet: Garment generation and self-supervised draping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Tao Du, Kui Wu, Pingchuan Ma, Sebastien Wah, Andrew Spielberg, Daniela Rus, and Wojciech Matusik. Diffpd: Differentiable projective dynamics. *ACM Transactions on Graphics (TOG)*, 41 (2):1–21, 2021.

C Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. Brax–a differentiable physics engine for large scale rigid body simulation. *arXiv preprint arXiv:2106.13281*, 2021.

Artur Grigorev, Michael J Black, and Otmar Hilliges. Hood: Hierarchical graphs for generalized modelling of clothing dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16965–16974, 2023.

Kai He, Kaixin Yao, Qixuan Zhang, Jingyi Yu, Lingjie Liu, and Lan Xu. Dresscode: Autoregressively sewing and generating garments from text guidance. *ACM Transactions on Graphics (TOG)*, 43(4):1–13, 2024.

Philipp Herholz, Tuur Stuyck, and Ladislav Kavan. A mesh-based simulation framework using automatic code generation. *ACM Transactions on Graphics (TOG)*, 43(6):1–17, 2024.

Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. Difftaichi: Differentiable programming for physical simulation. *arXiv preprint arXiv:1910.00935*, 2019a.

Yuanming Hu, Jiancheng Liu, Andrew Spielberg, Joshua B Tenenbaum, William T Freeman, Jiajun Wu, Daniela Rus, and Wojciech Matusik. Chainqueen: A real-time differentiable physical simulator for soft robotics. In *2019 International conference on robotics and automation (ICRA)*, pp. 6265–6271. IEEE, 2019b.

Zizhou Huang, Davi Colli Tozoni, Arvi Gjoka, Zachary Ferguson, Teseo Schneider, Daniele Panozzo, and Denis Zorin. Differentiable solver for time-dependent deformation problems with contact. *ACM Transactions on Graphics*, 43(3), 2024.

Krishna Murthy Jatavallabhula, Miles Macklin, Florian Golemo, Vikram Voleti, Linda Petrini, Martin Weiss, Breandan Considine, Jérôme Parent-Lévesque, Kevin Xie, Kenny Erleben, et al. gradsim: Differentiable simulation for system identification and visuomotor control. *International Conference on Learning Representations (ICLR)*, 2021.

Boyi Jiang, Juyong Zhang, Yang Hong, Jinhao Luo, Ligang Liu, and Hujun Bao. Bcnet: Learning body and cloth shape from a single image. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, pp. 18–35. Springer, 2020.

Maria Korosteleva and Sung-Hee Lee. Neuraltailor: Reconstructing sewing pattern structures from 3d point clouds of garments. *ACM Transactions on Graphics (TOG)*, 41(4):1–16, 2022.

Maria Korosteleva and Olga Sorkine-Hornung. Garmentcode: Programming parametric sewing patterns. *ACM Transactions on Graphics (TOG)*, 42(6):1–15, 2023.

Boqian Li, Haiwen Feng, Zeyu Cai, Michael J. Black, and Yuliang Xiu. ETCH: Generalizing Body Fitting to Clothed Humans via Equivariant Tightness. In *ICCV*, October 2025a.

Minchen Li, Zachary Ferguson, Teseo Schneider, Timothy R Langlois, Denis Zorin, Daniele Panozzo, Chenfanfu Jiang, and Danny M Kaufman. Incremental potential contact: intersection- and inversion-free, large-deformation dynamics. *ACM Trans. Graph.*, 39(4):49, 2020.

Minchen Li, Danny M. Kaufman, and Chenfanfu Jiang. Codimensional incremental potential contact. *ACM Trans. Graph. (SIGGRAPH)*, 40(4), 2021.

Ren Li, Corentin Dumery, Benoît Guillard, and Pascal Fua. Garment recovery with shape and deformation priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1586–1595, 2024a.

Ren Li, Cong Cao, Corentin Dumery, Yingxuan You, Hao Li, and Pascal Fua. Single view garment reconstruction using diffusion mapping via pattern coordinates. *ACM SIGGRAPH*, 2025b.

Xuan Li, Chang Yu, Wenxin Du, Ying Jiang, Tianyi Xie, Yunuo Chen, Yin Yang, and Chenfanfu Jiang. Dress-1-to-3: Single image to simulation-ready 3d outfit with diffusion prior and differentiable physics. *arXiv preprint arXiv:2502.03449*, 2025c.

Yifei Li, Tao Du, Kui Wu, Jie Xu, and Wojciech Matusik. Diffcloth: Differentiable cloth simulation with dry frictional contact. *ACM Transactions on Graphics (TOG)*, 42(1):1–20, 2022.

Yifei Li, Hsiao-yu Chen, Egor Larionov, Nikolaos Sarafianos, Wojciech Matusik, and Tuur Stuyck. Diffavatar: Simulation-ready garment optimization with differentiable simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4368–4378, 2024b.

Junbang Liang, Ming Lin, and Vladlen Koltun. Differentiable cloth simulation for inverse problems. *Advances in Neural Information Processing Systems*, 32, 2019.

Lijuan Liu, Xiangyu Xu, Zhijie Lin, Jiabin Liang, and Shuicheng Yan. Towards garment sewing pattern reconstruction from a single image. *ACM Transactions on Graphics (TOG)*, 42(6):1–15, 2023.

Shengqi Liu, Yuhao Cheng, Zhuo Chen, Xingyu Ren, Wenhan Zhu, Lincheng Li, Mengxiao Bi, Xiaokang Yang, and Yichao Yan. Multimodal latent diffusion model for complex sewing pattern generation. *arXiv preprint arXiv:2412.14453*, 2024.

Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1096–1104, 2016.

Zhongjin Luo, Haolin Liu, Chenghong Li, Wanghao Du, Zirong Jin, Wanhu Sun, Yinyu Nie, Weikai Chen, and Xiaoguang Han. Garverselod: High-fidelity 3d garment reconstruction from a single in-the-wild image using a dataset with levels of details. *ACM Transactions on Graphics (TOG)*, 43(6):1–12, 2024.

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

Miles Macklin, Matthias Müller, and Nuttapong Chentanez. Xpbd: position-based simulation of compliant constrained dynamics. In *Proceedings of the 9th International Conference on Motion in Games*, pp. 49–54, 2016.

Charles C Margossian. A review of automatic differentiation and its efficient implementation. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 9(4):e1305, 2019.

Gyeongsik Moon, Hyeongjin Nam, Takaaki Shiratori, and Kyoung Mu Lee. 3d clothed human reconstruction in the wild. In *European conference on computer vision*, pp. 184–200. Springer, 2022.

Matthias Müller, Bruno Heidelberger, Marcus Hennix, and John Ratcliff. Position based dynamics. *Journal of Visual Communication and Image Representation*, 18(2):109–118, 2007.

Kiyohiro Nakayama, Jan Ackermann, Timur Levent Kesdogan, Yang Zheng, Maria Korosteleva, Olga Sorkine-Hornung, Leonidas J Guibas, Guandao Yang, and Gordon Wetzstein. Aipparel: A large multimodal generative model for digital garments. *arXiv preprint arXiv:2412.03937*, 2024.

Uwe Naumann. *The art of differentiating computer programs: an introduction to algorithmic differentiation*. SIAM, 2011.

Shuliang Ning, Duomin Wang, Yipeng Qin, Zirong Jin, Baoyuan Wang, and Xiaoguang Han. Picture: Photorealistic virtual try-on from unconstrained designs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6976–6985, 2024.

Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020.

Li Ren, Benoit Guillard, and Pascal Fua. Isp: Multi-layered garment draping with implicit sewing patterns. *Advances in Neural Information Processing Systems*, 2023.

Igor Santesteban, Miguel A Otaduy, and Dan Casas. Snug: self-supervised neural dynamic garments. ieee. In *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pp. 9, 2022.

Zehong Shen, Huaijin Pi, Yan Xia, Zhi Cen, Sida Peng, Zechen Hu, Hujun Bao, Ruizhen Hu, and Xiaowei Zhou. World-grounded human motion recovery via gravity-view coordinates. In *SIGGRAPH Asia Conference Proceedings*, 2024.

Tuur Stuyck and Hsiao-yu Chen. Diffxpbd: Differentiable position-based simulation of compliant constraint dynamics. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 6(3):1–14, 2023.

Tencent Hunyuan3D Team. Hunyuan3d 1.0: A unified framework for text-to-3d and image-to-3d generation, 2024.

Tencent Hunyuan3D Team. Hunyuan3d 2.0: Scaling diffusion models for high resolution textured 3d assets generation, 2025a.

Tencent Hunyuan3D Team. Hunyuan3d 2.5: Towards high-fidelity 3d assets generation with ultimate details, 2025b. URL https://arxiv.org/abs/2506.16504.

Haoyu Wang, Zhilu Zhang, Donglin Di, Shiliang Zhang, and Wangmeng Zuo. Mv-vton: Multi-view virtual try-on with diffusion models. *arXiv preprint arXiv:2404.17364*, 2024a.

Tuanfeng Y Wang, Duygu Ceylan, Jovan Popovic, and Niloy J Mitra. Learning a shared shape space for multimodal garment design. *ACM SIGGRAPH*, 2018.

Wenbo Wang, Hsuan-I Ho, Chen Guo, Boxiang Rong, Artur Grigorev, Jie Song, Juan Jose Zarate, and Otmar Hilliges. 4d-dress: A 4d dataset of real-world human clothing with semantic annotations. In *CVPR*, 2024b.

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34, 2021.

Yuliang Xiu, Jinlong Yang, Dimitrios Tzionas, and Michael J Black. Icon: Implicit clothed humans obtained from normals. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13286–13296. IEEE, 2022.

Yuliang Xiu, Jinlong Yang, Xu Cao, Dimitrios Tzionas, and Michael J Black. Econ: Explicit clothed humans optimized via normal integration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 512–523, 2023.

Jie Xu, Tao Chen, Lara Zlokapa, Michael Foshey, Wojciech Matusik, Shinjiro Sueda, and Pulkit Agrawal. An end-to-end differentiable framework for contact-aware robot design. *RSS*, 2021.

Shan Yang, Zherong Pan, Tanya Amert, Ke Wang, Licheng Yu, Tamara Berg, and Ming C Lin. Physics-inspired garment recovery from a single-view image. *ACM Transactions on Graphics (TOG)*, 37(5):1–14, 2018.

Boyang Yu, Frederic Cordier, and Hyewon Seo. Inverse garment and pattern modeling with a differentiable simulator. In *Computer Graphics Forum*, volume 43, pp. e15249. Wiley Online Library, 2024.

Fang Zhao, Wenhao Wang, Shengcai Liao, and Ling Shao. Learning anchored unsigned distance functions with gradient direction alignment for single-view garment reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12674–12683, 2021.

Feng Zhou, Ruiyang Liu, Chen Liu, Gaofeng He, Yong-Lu Li, Xiaogang Jin, and Huamin Wang. Design2garmentcode: Turning design concepts to tangible garments through program synthesis. *CVPR*, 2025.

Heming Zhu, Yu Cao, Hang Jin, Weikai Chen, Dong Du, Zhangye Wang, Shuguang Cui, and Xiaoguang Han. Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pp. 512–530. Springer, 2020.