

Stitched Embeddings: A Unified Latent Space for 3D Garments and 2D Patterns

Andrea Sanchietti^{1,2}, Riccardo Marin^{3,4}, Bharat Lal Bhatnagar⁶
Yuanlu Xu⁶, and Gerard Pons-Moll^{1,2,5}

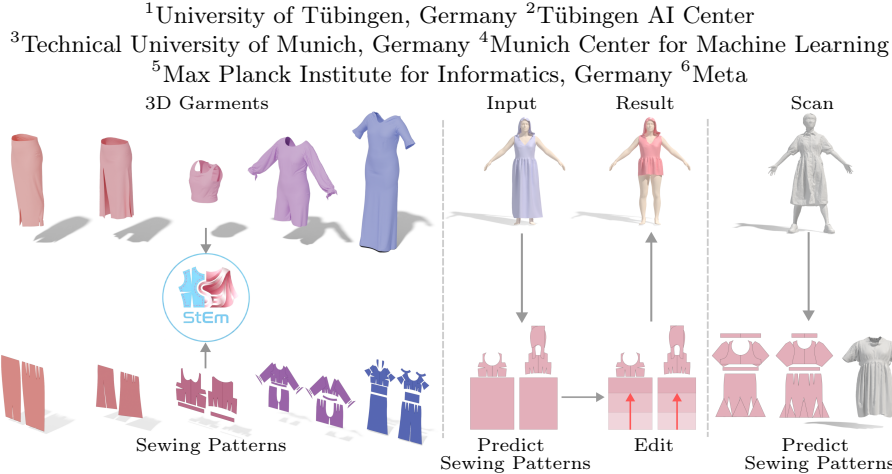


Fig. 1: In this work, we propose **Stitched Embeddings (StEm)**, a unified latent space to jointly represent 3D garments and their corresponding 2D sewing patterns. Our method, StEm-Net, provides state-of-the-art performance for sewing pattern prediction from 3D garments. Our approach is trained end-to-end, and supports test-time adaptation for sewing pattern predictions and seamless propagation of editing from sewing patterns to 3D realizations.

Abstract. While garments are essential for realistic digital humans, their topological variety makes them much harder to model than parametric bodies. Traditional tailoring relies on 2D sewing patterns, yet bridging these patterns to 3D geometry currently requires physical simulations. We present Stitched Embeddings, the first simulation-free framework to unify 3D garment reconstruction and sewing pattern inference within a single bidirectional latent space. By leveraging the geometric priors of a pretrained 3D foundation model, our approach overcomes the data scarcity typically associated with high-quality garment modeling. We propose to use the BoxMesh as a critical intermediate representation to align 2D panels into 3D configurations without the computational overhead of a simulator. This architecture achieves state-of-the-art accuracy in pattern reconstruction while significantly improving efficiency. Furthermore, our differentiable pipeline enables novel applications, including pattern recovery from meshes and 3D editing from 2D patterns. Finally, this work provides a scalable link between neural 3D vision and the physical garment manufacturing pipeline. Project page: <https://andreas00.github.io/stitchedembeddings>.

1 Introduction

The rapid progress of 3D capture technology, together with advances in data-driven techniques, has opened unprecedented opportunities for human digitalization. Within this domain, garments constitute a fundamental pillar of individual identity and visual realism. But if Virtual Human research has provided dramatic advances thanks to the development of parametric body models, faithful garment reconstruction and manipulation have remained limited and far from being solved. The vast topological variability and non-rigid deformation of clothing make it mathematically intractable to reduce garments to a simple additive mesh model similar to SMPL [39]. Hence, developments in neural fields have motivated scholars to represent them as unsigned distance fields (UDFs) [9, 10, 20, 36, 56].

However, implicit representations lack the semantic structure required for downstream applications in the tailoring and fashion industries. Manufacturing needs sewing patterns: a 2D, piece-wise CAD formalization that serves as the blueprint for production and realistic animations. Formally, these patterns consist of discrete 2D panels whose boundaries are defined by parametric curves (*e.g.*, Bézier curves), interconnected via seam constraints. Traditionally, the bridge between 2D patterns and 3D geometry is established through physics-based simulation, in which panels are initialized in 3D space and "sewn" together using spring-mass systems. Researchers have collected several datasets [1, 26, 28], defined suitable tokenizations [19, 46], and achieved promising results in the prediction of sewing patterns from images [5, 6] or point clouds [28].

Despite all these efforts, we still observe a lack of scalability and generalization in the methods. Recent approaches aim to predict sewing patterns from 3D using a single-pass regression, followed by a physical simulation to obtain the final 3D geometry. However, physical simulators are slow and generally not differentiable, which prevents end-to-end learning and their use for optimization (*e.g.*, recovery from incorrect predictions through test-time adaptation). More recently, differentiable physical simulators have become more affordable [23, 41, 73]. However, they remain slow and still do not support editing the number or type of sewing patterns during test-time optimization. Our research originates from a fundamental question: how can we combine the flexibility of the UDF representation with the controllability and application relevance of the sewing pattern?

Our answer is **Stitched Embeddings**: a common latent space where 3D Garments and 2D Sewing Pattern Parameters coexist. We build such a space using a combination of 3D garments and BoxMesh, a representation in which sewing patterns are arranged in 3D as an initialization for a simulation. Our intuition is that it can serve as a spatial proxy that aligns 2D panels into a canonical 3D configuration according to their semantic function (*e.g.* sleeves, front panel). By combining their spatial locations with their semantic meanings, we significantly facilitate learning a common latent space (see Fig. 1 for recovered samples from such a latent space). Our Stitched Embedding Network (StEm-Net) is built on top of a 3D foundational model, demonstrating for the first time the use of such backbones for 2D sewing patterns.

We measure our performance on two datasets of different complexity, showing that StEm-Net achieves state-of-the-art performance in sewing pattern prediction from 3D point clouds, improving metrics even up to 15% over the closest competitor. Remarkably, although extracting semantics from 3D point clouds is more challenging than from 2D images, our method also surpasses 2D-based methods. In addition, the differentiability of our approach enables test-time latent optimization. We can further improve the network’s initial prediction by backpropagating a reconstruction loss to yield sewing patterns that better align with the input 3D garment. Also, StEm allows us to edit one representation and observe the impact of the modifications on the other. For example, we can update the sewing pattern and obtain an updated 3D garment within the same framework, without requiring backpropagation or physical simulation. Ultimately, this work represents a significant milestone towards a scalable, end-to-end learning framework that bridges the gap between neural 3D vision and the physical tailoring pipeline.

In summary, our contributions are:

1. *StEm-Net*: Our Stitched Embedding Network (StEm-Net) is the first end-to-end, fully differentiable, and simulation-free method that connects different garment representations in a unified latent space.
2. We set a new state of the art in the task of sewing pattern prediction from 3D point clouds, reducing the error by up to 15% compared to the closest competitor.
3. We enable test-time adaptation and direct garment editing within the same framework, without the need for physical simulations, providing a practically flexible and useful tool for the Vision and Graphics communities.

2 Related Works

2.1 Garment Reconstruction

Reconstructing 3D garments from disparate inputs (images, videos, noisy scans) has attracted broad interest [13, 36, 42, 49, 59, 67], motivating significant efforts in dataset collections and annotation [1, 14, 21, 26, 43–45, 60, 79], including procedural approaches [27]. Despite the growing availability of such data, the space of possible garment shapes and deformations is vast and far from being fully covered. Hence, many methods reconstruct by relying on a fixed template for clothing [21, 75] or a parametric template for the human body [1, 4, 33, 39, 54, 55]. On the other hand, such templates also constrain the garments that can be represented. Another characteristic of all the aforementioned methods is their use of Unsigned Distance Fields (UDF) as the main output for garments. Such a representation is convenient, as it allows flexible topology and continuous deformation. Other works combine 3D meshes and UDF fields to generate or reconstruct continuous surfaces, such as SurfD [74], Ghost-on-the-Shell [38], GarmentCrafter [61], and SimAvatar [71]. Recent works also find applications of Gaussian Splatting [15, 18, 25, 37, 50, 58], while adapting them to practical use

cases (*e.g.*, modeling and animation) remains an open research area [62,63,66,77]. In this paper, we do not assume a fixed template or mesh, and instead propose connecting the output UDF to a parametric representation of garments, building on recent advances in sewing pattern modeling.

2.2 Sewing Pattern Recovery

From Single Image. In recent years, we have witnessed significant work on providing garments with an underlying parametric structure, also called sewing patterns [40, 48, 51, 64, 65] (see Fig. 2). Such a representation is borrowed from standard garment manufacturing, where clothes are obtained by designing small patches that are then sewn together. Sewing patterns are advantageous for applications and for bridging the gap between research and industry. Also, they provide a piecewise regularization, flexible enough to represent a vast set of garments, and also enforce physical plausibility. Since the sewing patterns naturally live on a 2D plane, a line of work tries to recover them and the 3D garments from such a flat representation [72]. The work of Berthouzoz et al. [3] was the first to propose a parsing from sewing pattern diagrams to 3D garment models, relying on a combination of machine learning and integer programming. More recently, Panelformer [7] and SewFormer [34] proposed reconstructing patterns from images by extracting features with a transformer and predicting sewing pattern panels and edges. Recent works such as GarmentX and DressCode [17, 19] have explored autoregressive architectures to generate garment parameters from images. ChatGarment [5] and AIpparel [46] finetune VLMs [76] to generate patterns from text or image inputs. Design2GarmentCode [78] tackles the problem of predicting sewing patterns as a code-generation problem. GarmentImage [52] uses a raster-based sewing pattern representation. SewingLDM [35] and GarmentDiffusion [32] explore methods to encode patterns in a compact latent space and then learn a diffusion process conditioned on input text or sketches. Though this line of work has been promising due to the limited input required, it is also highly underconstrained. With advances in acquisition and generation techniques, the availability of 3D models has become increasingly common [11, 12, 68–70]. This motivated several methods to start directly from 3D assets.

From 3D. With the increased availability of 3D data, we witness an emergent attention toward recovering sewing patterns from 3D clothes. While early works were restricted to modeling simple garments such as skirts and t-shirts [2, 8], the more recent NeuralTailor [28] reconstructs sewing patterns from 3D point clouds and trains on the GarmentData [27] dataset, which has a lower number of patterns than the GarmentCodeData [26, 29], and predicts patterns by regressing edges and stitches of panels with LSTMs [22]. The main downside of NeuralTailor is its reliance on a final garment mesh simulation, which hinders the possibility of backpropagating through the network to optimize the pattern predictions. GarmentNet [31] introduces a unified 2D–3D garment representation that encodes each sewing-pattern panel as a geometry image [16] in UV space, jointly capturing panel contours and draped 3D geometry. These panel-wise representations

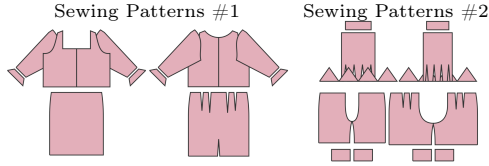


Fig. 2: An example of sewing patterns. Every garment is represented by a set of panels. While sometimes their semantic meaning is clear (*e.g.*, for the body), for others it is difficult to guess without simulating them (*e.g.*, the pieces on the right are a skirt and pants).

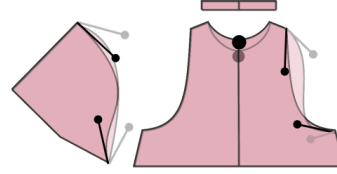


Fig. 3: An example of Sewing Pattern variability. Every piece is defined by Bézier curves; their shape can be modified by acting on the control points.

are compressed into a latent space and generated jointly using a diffusion transformer. Their method enables pattern-based reconstruction of garments from images and point clouds. In a concurrent work, named Garment Particles [47], authors propose to diffuse a 5D point cloud that jointly models 3D geometry and 2D patterns. Another concurrent work Inverse Draping [24] proposes formulating the pattern prediction from meshes as an autoregressive BoxMesh generation. Finally, ReWeaver [30] is a contemporary work that starts from multi-views, uses VGGT [57] to recover a point cloud, and then predicts the SP parameters. None of the previous methods provides an end-to-end approach to go from 3D to a sewing pattern and vice versa, and they require complicated machinery to combine 3D and sewing patterns. This makes it impossible to run any optimization on top of them.

3 Background

3.1 Sewing Pattern Representation

In traditional garment modeling, sewing patterns are defined as a collection of 2D boundary curves (typically a mixture of linear segments and quadratic or cubic Bézier curves, see Figs. 2 and 3) representing flat fabric panels. Formally, a pattern can be described by a set of vertices $V \subset \mathbb{R}^2$, edges E , and a set of curvature parameters C that define the spline geometry. While this representation is precise for manufacturing, it is high-dimensional, non-uniform in length, and difficult for standard regression networks to predict directly.

To overcome this, we leverage the parametric space of GarmentCode [29]. Instead of predicting individual Bézier control points, our model operates on a fixed-dimensional set of **design parameters** (122 for GarmentCodeData [26]). These parameters provide a high-level, low-dimensional abstraction of the garment (*e.g.* sleeve length, waist width, or neckline style) from which the full geometric sewing patterns can be deterministically generated.

To bridge the gap between 2D patterns and 3D geometry, we utilize a proxy representation named the **BoxMesh**. The BoxMesh is generated using GarmentCode, which places the 2D patterns in 3D space around a body model and applies a preliminary meshing procedure. It maintains the exact topology of the sewing patterns but occupies a 3D volume. When a cloth simulation is applied

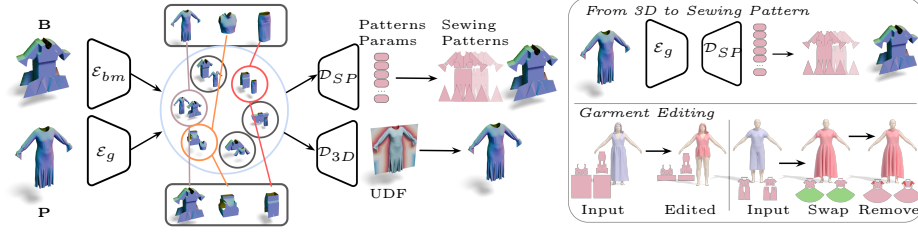


Fig. 4: On the left: StEm-Net is the first end-to-end differentiable method that unifies multiple different representations in a common compact representation. The model can map BoxMeshes and 3D garments into Stitched Embeddings. From such a latent space, we decode sewing pattern parameters and UDFs. On the right, we show our method’s modalities: prediction of Pattern Parameters and UDFs from an input Mesh (top), and editing of 3D Garments based on 2D edits (bottom).

to the BoxMesh, it yields the final deformed **Garment Mesh**. By using the BoxMesh as an input modality, we provide our network with 3D spatial priors that are topologically aligned with the final output.

3.2 Hunyuan3D’s VAE Architecture

Our architecture is built on top of the Hunyuan3D 2.1 [53] VAE framework, which we modify to better suit our task. While the original work utilizes Signed Distance Fields (SDF) to model watertight objects, we adopt an **Unsigned Distance Field (UDF)**. This allows our model to represent garments as non-watertight, open manifolds, which is essential for capturing realistic clothing structures like necklines and open hems.

A critical challenge in our cross-modal task is ensuring that the latent space is invariant to whether the input is a BoxMesh or a simulated Garment Mesh. Since the encoder processes meshes by sampling point clouds via Farthest Point Sampling (FPS), the resulting latent sets are inherently unordered. To ensure feature alignment, we exploit the fact that the BoxMesh and the Garment Mesh share an identical topology. During training, we sample **corresponding points** from both geometries. This ensures that the same physical locations on a garment, whether in its pre-simulated “box” state or its final draped state, map to the same regions in the latent space.

4 Stitched Embedding Networks (StEm-Net)

Overview. In this section, we describe StEm-Net, the first learning-based approach differentiable end-to-end, that unifies 3D garment realizations and sewing patterns in a common shared latent space. We call such a representation Stitched Embeddings (StEm), and in the following, we describe how we design its encoding and decoding. Next, we outline the training procedure of StEm-Net. We present our pipeline in Fig. 4.

StEm encoding. Our goal is to define a common, grounding representation, where UDFs and sewing patterns are paired. Such a task is far from trivial: UDFs are 3D continuous representations, well modeled by parametric functions such as neural fields. Sewing Patterns are a discrete set of 2D pieces, parametrized by a continuous boundary. To date, we are unaware of any method capable of providing a shared latent space for these two different representations. To solve this, we propose Stitched Embeddings: a latent space that combines 3D garments with the BoxMesh of their sewing patterns, providing spatial connections to the 3D UDFs while also expressing the semantics of the 2D sewing patterns. To do so, we rely on two encoders: one for the 3D garments ($\mathcal{E}_g$) and one for the pre-simulation BoxMesh ($\mathcal{E}_{bm}$). The garment encoder $\mathcal{E}_g$ takes as input a garment represented as a point cloud $\mathbf{P} \in \mathbb{R}^{5120 \times 3}$ equipped with normals $\mathbf{N} \in \mathbb{R}^{5120 \times 3}$. The BoxMesh encoder $\mathcal{E}_{bm}$ processes the pre-simulated geometry. At training time, in order to align the latent distribution of the BoxMesh encoder with that of the garment encoder, we exploit the known 1 : 1 correspondence between the simulated garment and its pre-simulation BoxMesh. Instead of independently sampling points from the BoxMesh surface, we construct a point cloud $\mathbf{B} \in \mathbb{R}^{5120 \times 3}$ such that each point in $\mathbf{B}$ corresponds exactly to a point in $\mathbf{P}$. To ensure consistency of the input representation, we associate with $\mathbf{B}$ the same normals $\mathbf{N}$ used for the garments. In our implementation, both encoders are inherited from Hunyuan3D [53] and map to a common latent set $\mathbf{z} \in \mathbb{R}^{256 \times 64}$.

StEm decoding. We are interested in two representations: the UDF of the 3D garments positioned in space, and the related sewing patterns. These two are the most useful in practice: the UDF provides a 3D continuous representation that can be optimized to fit observations such as scans, while sewing patterns are fundamental for manufacturing (*e.g.*, garment editing) and graphics pipelines (*e.g.*, animation). Hence, the 3D decoder $\mathcal{D}_{3D}$ reconstructs the UDF field (truncated at 0.01), while the sewing pattern decoder $\mathcal{D}_{SP}$ regresses the 122 design parameters. We decode these parameters from the latent set with a decoding transformer, after which learnable per-panel queries cross-attend over the full latent sequence, yielding one feature per panel that lightweight MLP heads map to that panel’s parameters. These parameters correspond exactly to the parametric design space defined by the underlying garment modeling system [29], which deterministically generates sewing patterns from these structured parameters. Consequently, our prediction space aligns with the expressive range of the parametric generator, ensuring valid, manufacturable sewing patterns by construction. To ensure the model remains body-aware, all meshes are normalized relative to the transformation required to fit the underlying body model into a $[-1, 1]$ unit cube.

4.1 StEm-Net Training Losses.

Our architecture requires the model both to produce reliable reconstructions and to learn a latent representation in which sewing patterns and 3D models of the same garment are close together. Hence, we optimize StEm-Net with a

composite loss function:

$$\mathcal{L} = \mathcal{L}_{UDF} + \lambda_1 \mathcal{L}_{SP} + \lambda_2 \mathcal{L}_{KL} \quad (1)$$

Where $\mathcal{L}_{UDF}$ is a 3D reconstruction loss, $\mathcal{L}_{SP}$ is a sewing pattern loss, and $\mathcal{L}_{KL}$ is used to enforce latent alignment.

3D Reconstruction Loss The UDF loss minimizes the Mean Squared Error (MSE) between the predicted and ground truth distance fields for both garment and BoxMesh inputs:

$$\mathcal{L}_{UDF} = \sum_{p \in \Omega} \text{MSE}(\mathcal{D}_{3D}(\mathcal{E}_g(\mathbf{P}, \mathbf{N}), p), \text{UDF}(p)) \quad (2)$$

$$+ \sum_{p \in \Omega} \text{MSE}(\mathcal{D}_{3D}(\mathcal{E}_{bm}(\mathbf{B}, \mathbf{N}), p), \text{UDF}(p)) \quad (3)$$

where p represents points sampled within the volume $\Omega = [-1, 1]^3 \subset \mathbb{R}^3$, and $\text{UDF}(p)$ returns the ground truth distance of p from the surface. We adopt a stratified sampling strategy, by sampling half of the points randomly, and half at varying distances (*i.e.*, ± 0.0015 , ± 0.005 , ± 0.025 , ± 0.05) from the surface to capture both high-frequency details and the global field structure.

Sewing Pattern Loss The 122 design parameters are heterogeneous, comprising continuous, boolean, and categorical values. We therefore assign each parameter to a prediction head, and split $\mathcal{L}_{SP}$ into:

$$\mathcal{L}_{SP} = \sum_{i \in \text{cont}} \text{MSE}(y_i, \hat{y}_i) + \sum_{j \in \text{bool}} \text{BCE}(y_j, \hat{y}_j) + \sum_{k \in \text{cat}} \text{CE}(y_k, \hat{y}_k) \quad (4)$$

where MSE handles continuous dimensions (*e.g.* panel dimensions), BCE handles binary flags (*e.g.* sleeve presence), CE handles multi-class categories (*e.g.*, garment type), y_n represents the output of the n -th prediction head, and $\hat{y}_n$ represents the ground truth value of the n -th parameter.

Latent Alignment The KL-divergence loss serves a dual purpose: regularizing the latent space against a standard normal distribution and enforcing consistency between the two input modalities:

$$\mathcal{L}_{KL} = \text{KL}(z_G \parallel \mathcal{N}(0, I)) + \text{KL}(z_B \parallel z_G) \quad (5)$$

This formulation forces the BoxMesh encoding z_B to reside in the same manifold as the garment encoding z_G , allowing the model to infer 2D patterns from 3D scans effectively.

4.2 StEm-Net Inference

StEm-Net enables bidirectional mapping between the 3D geometric domain and the 2D sewing pattern space.

From 3D to Sewing Patterns For inference from 3D observations, we encode a point cloud $\mathbf{P}$ and normals $\mathbf{N}$ through $\mathcal{E}_g$ to obtain its representation in

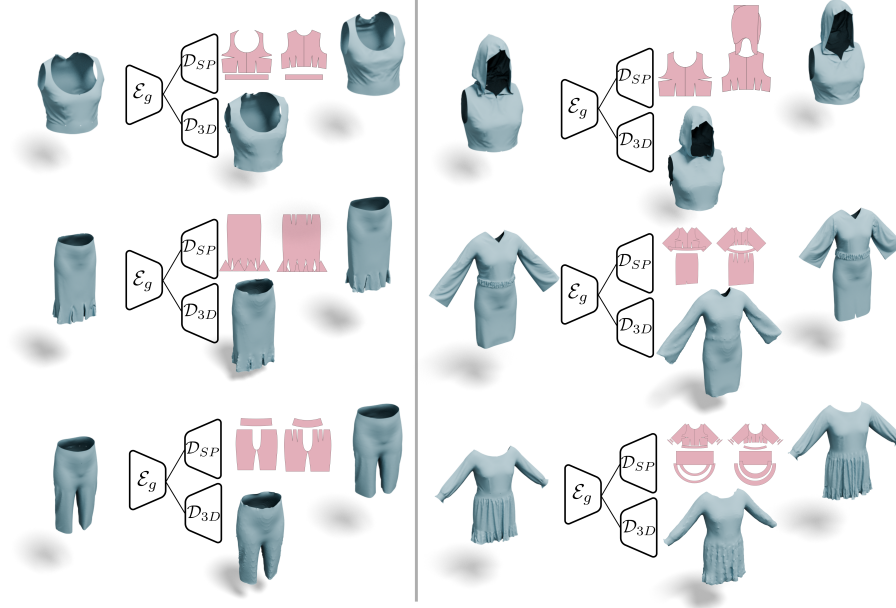


Fig. 5: Autoencoding of StEm-Net. On the left, we show the input mesh. We pass this mesh to our garment encoder $\mathcal{E}_g$, which maps it into our StEm latent space. From that, we recover the 3D garment via $\mathcal{D}_{3D}$, and the sewing patterns via $\mathcal{D}_{SP}$, which we also simulate to obtain the corresponding 3D. We can appreciate that both the 3D output and the simulated sewing patterns produce geometry close to the ground truth, proving the reliability of our approach.

the StEm latent space. From this shared embedding, the decoder $\mathcal{D}_{SP}$ regresses structured design parameters, reconstructing the 2D manufacturing specifications directly from the 3D geometry. While our encoders provide a robust initialization, the mapping from unseen 3D geometries to the latent space is not always lossless. To bridge this domain gap and refine the predicted sewing patterns, we perform *test-time latent optimization*. Given an input mesh, we initialize its latent vector set $\mathbf{z}$ and iteratively refine it by minimizing the reconstruction error between the predicted UDF and the input surface:

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \sum_{p \in \Omega} \text{MSE}(\mathcal{D}_{3D}(\mathbf{z}, p), \text{UDF}(p)) \quad (6)$$

where $\Omega \subset \mathbb{R}^3$ represents the set of points sampled within the bounding volume, $\mathcal{D}_{3D}(\mathbf{z}, p)$ is the distance predicted by the 3D decoder at point p given the latent set $\mathbf{z}$ and $\text{UDF}(p)$ is the ground truth distance from the input surface.

StEm is the first approach whose test-time optimization can modify both the continuous geometry and the discrete structure of a sewing pattern. Because the StEm latent space is shared and differentiable, the gradient-based refinement of the 3D geometry directly propagates to the sewing pattern parameters. This allows the model to “correct” the 2D design in response to the 3D observations,

Method	NeuralTailor Testset					GarmentCodeData TestSet				
	Panel L2 ↓	#Panels	#Edges	Stitch Precision	Stitch Recall	Panel L2 ↓	#Panels	#Edges	Stitch Precision	Stitch Recall
NeuralTailor [28]	5.2	83.6%	87.3%	74.7%	83.9%	-	-	-	-	-
StEm-Net	4.6	81.6%	93.3%	85.7%	86.0%	4.4	69.5%	82.6%	66.6%	66.4%
StEm-Net + Optim	4.4	84.6%	95.0%	88.5%	89.2%	4.3	71.1%	80.6%	64.0%	64.3%

Table 1: Comparison of NeuralTailor and GarmentCodeData test sets. On NeuralTailor, our method provides better performance both before and after the optimization step. On GarmentCodeData, StEm-Net provides a similar Panel L2 error, while NeuralTailor fails to converge.

ensuring the final output is not just a static prediction but an optimized fit. Our experiments demonstrate that this refinement improves the accuracy of the regressed design parameters, particularly for out-of-distribution garments.

In Fig. 5, we show the result of autoencoding the input mesh while predicting the pattern. To validate these results, we computed the mean Chamfer Distance between the output of $\mathcal{D}_{3D}$ and the simulation result obtained from $\mathcal{D}_{SP}$ ’s output, yielding a value of 5.5 cm.

From Sewing Patterns to 3D Because the mapping from a 2D pattern to 3D is inherently one-to-many (as garments can drape in various states), models trained only on geometry often produce over-smoothed results. To resolve this ambiguity, we condition our encoding on surface normals $\mathbf{N}$. These directional priors transform the ill-posed $1 : N$ mapping into a deterministic $1 : 1$ correspondence, allowing the network to capture high-frequency folds and generalize to unseen drapes. StEm-Net also functions as a high-speed “neural simulator”. By processing a pre-simulation BoxMesh through $\mathcal{E}_{bm}$, the model predicts the corresponding 3D UDF field in a fraction of a second, bypassing traditional physics-based solvers.

5 Results

Baselines. We evaluate our method against NeuralTailor [28], as it represents the most closely related approach to end-to-end sewing pattern estimation. NeuralTailor utilizes a PointNet-based encoder to process input scans and regresses pattern parameters through an LSTM-based decoder. To isolate the impact of our refinement process, we also include an ablation of our model without test-time optimization. While we attempted to train NeuralTailor on the more complex GarmentCodeData, the model failed to converge. We attribute this to the significant increase in structural complexity: GarmentCodeData [26] contains up to 75 panels per garment (compared to 24 in the original GarmentData) and features a variable number of edges per panel, whereas the NeuralTailor dataset assumes a fixed topology. Furthermore, the significantly larger scale of GarmentCodeData likely exceeded the representational capacity of the baseline’s sequential decoding architecture. Consequently, we report comparative results on the original GarmentData [28] to ensure a fair evaluation. For training and testing, we used the public train-test splits for the datasets.

GarmentCodeData TestSet					
Method	Panel L2 ↓	#Panels	#Edges	Stitch Precision	Stitch Recall
ChatGarment [5]	10.1	12.5%	45.5%	24.1%	25.1%
StEm-Net	4.4	69.5%	82.6%	66.6%	66.4%
StEm-Net + Optim	4.3	71.1%	80.6%	64.0%	64.3%

Table 2: Results on the GarmentCodeData TestSet. The optimization step slightly improves Panel L2 and panel detection, while stitch metrics remain comparable.

5.1 Comparison with 3D-based Baselines

We evaluate our method on the task of estimating sewing patterns from 3D point clouds. We compare StEm-Net against NeuralTailor [28], which, to date, is the only method with available code that solves an analogous task. As data, we consider the NeuralTailor and GarmentCode test sets, and in Tab. 1 we report our results. For the NeuralTailor, we used their pretrained checkpoint for comparison and trained our method on their training set. We observe that our method outperforms the baseline across almost all metrics, even before the optimization procedure. Particularly significant is the gain in the number of edges and stitch precision, demonstrating that our method is much more accurate in understanding the correct shapes of the sewing patterns and their interconnection. Likely, the use of BoxMesh provides better semantic supervision of individual patches during learning. After test-time optimization, we achieve even further improvements, with around 15% better L2 reconstruction and almost 15% better stitching precision. We also tried to compare ourselves on GarmentCode, but our attempts to train NeuralTailor on such a dataset failed to converge. Such a failure is likely caused by GarmentCode’s complexity. Our model, on the other hand, converges successfully. Hence, we provide the first method that actually works on GarmentCode, offering an operational baseline for this dataset. To provide a better intuition for the quality of our predictions, we present several qualitative results in Fig. 6. For each sample, we report the input 3D garment and the ground-truth sewing patterns on the left, while on the right, we show the sewing pattern predictions of the methods and the 3D representation generated by physical simulation. Our method achieves significantly more accurate results.

5.2 Comparison with Image-based Baselines

To evaluate the robustness of StEm-Net, we compare our performance against **ChatGarment** [5], a state-of-the-art method for regressing sewing patterns from 2D images, and we report the results in Tab. 2. While StEm-Net operates directly on 3D point clouds, ChatGarment utilizes a multi-modal approach. To ensure a fair comparison, we provide ChatGarment with 2D images where garment panels are explicitly segmented into distinct patches (as shown in Fig. 7). This comparison highlights a fundamental difference in input representation. ChatGarment relies on visual textures and 2D spatial arrangements to infer the pattern. In contrast, StEm-Net leverages 3D geometry and surface normals, which provide more definitive cues regarding physical draping and panel connectivity.

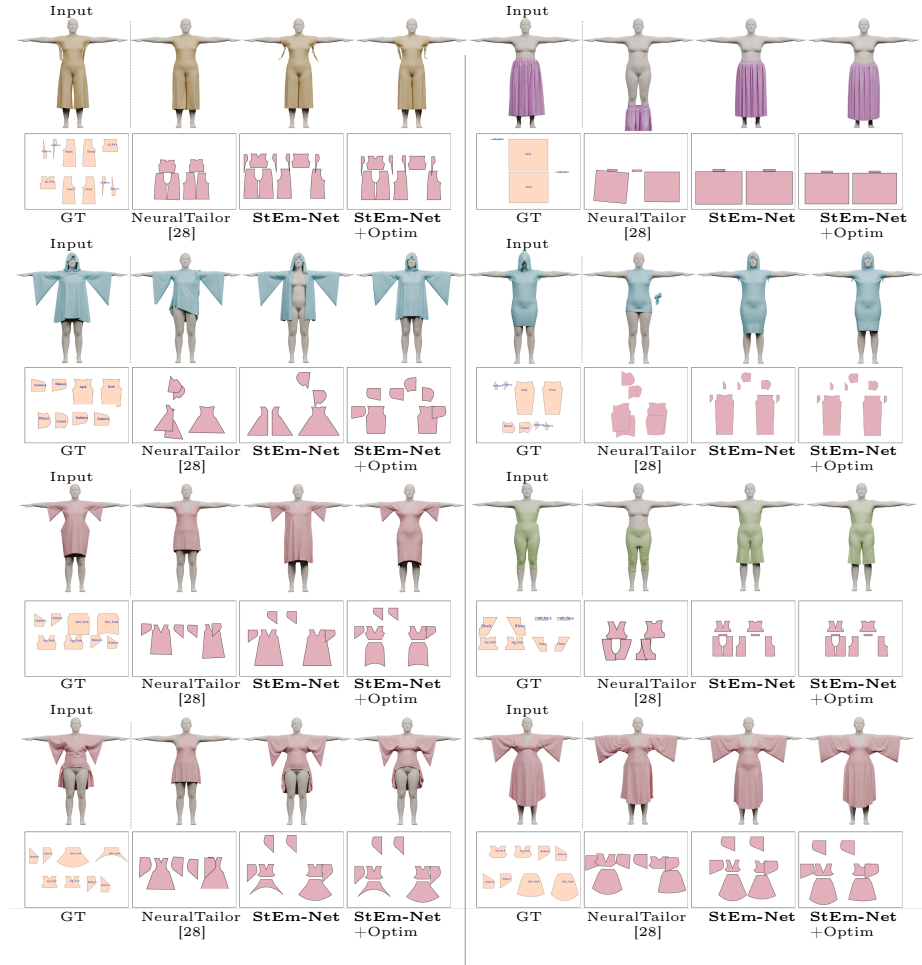


Fig. 6: Qualitative comparison on NeuralTailor test set. Our method provides robust predictions, even in the presence of nuances such as strings or hoodies with long, large sleeves. NeuralTailor [28] predicts inaccurate patterns, often resulting in dramatic failures. Our test-time latent optimization often recovers nuances and fixes inconsistencies, whereas in the worst case, it preserves the quality of the original prediction.

Even though the baseline benefits from pre-segmented patches, which significantly simplify topological inference, our experiments demonstrate that StEm-Net achieves superior accuracy in parameter regression. We attribute this to the StEm latent space, which preserves the structural relationship between 3D volume and 2D topology more effectively than a 2D-to-2D mapping. Furthermore, our test-time optimization allows StEm-Net to iteratively refine the pattern based on the input surface, a dynamic capability that purely feed-forward image-based methods lack.

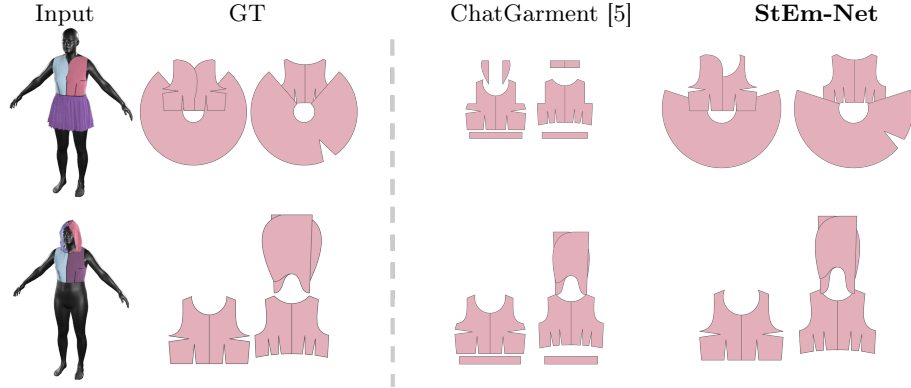


Fig. 7: Qualitative comparison on GarmentCodeData [26]. From left to right: input (images for ChatGarment [5] and 3D meshes for StEm-Net), ground truth patterns, and the respective predictions from ChatGarment and StEm-Net with test-time optimization.

5.3 Editing from patterns

A key advantage of our shared latent space is the ability to perform 3D garment editing via 2D pattern manipulation, without requiring a physical simulator. By initializing a garment in a specific configuration, users can explicitly modify the 2D sewing pattern parameters and observe the corresponding 3D drape through the forward pass of our neural simulator ($\mathcal{E}_{bm} \rightarrow \mathcal{D}_{3D}$). StEm-Net enables a seamless mapping from parametric 2D edits to high-fidelity 3D mesh updates. This provides a fast, physics-free alternative to traditional CAD pipelines, enabling intuitive, interactive garment design. During the edit, we assume the input normals do not change. Even if such an approximation produces sub-optimal input, our results show that the network is robust to it and operates normally. We report examples of our editing in Fig. 8.

5.4 Generalization to Real Scans

Although trained exclusively on curated synthetic garments, we evaluate our method’s ability to generalize to real-world scans—a highly challenging setting due to inherent sensor noise and topological inaccuracies. To process this data, we utilize scans from the 4DDress and CloSe datasets, specifically selecting instances with wide arm and leg stances. We first unpose the scans to a canonical A-pose and leverage the dataset’s segmentation labels to isolate individual garments before inference.

As shown in Fig. 9, our method accurately reconstructs 3D garments from raw scan data, demonstrating its practical viability for real-world digital fabrication and streamlining garment acquisition pipelines. Furthermore, we provide a quantitative comparison against NeuralTailor [28] in Table 3, measuring the Chamfer Distance between the predicted pattern and the scan on the CloSe samples. Our approach significantly outperforms the baseline across all metrics.

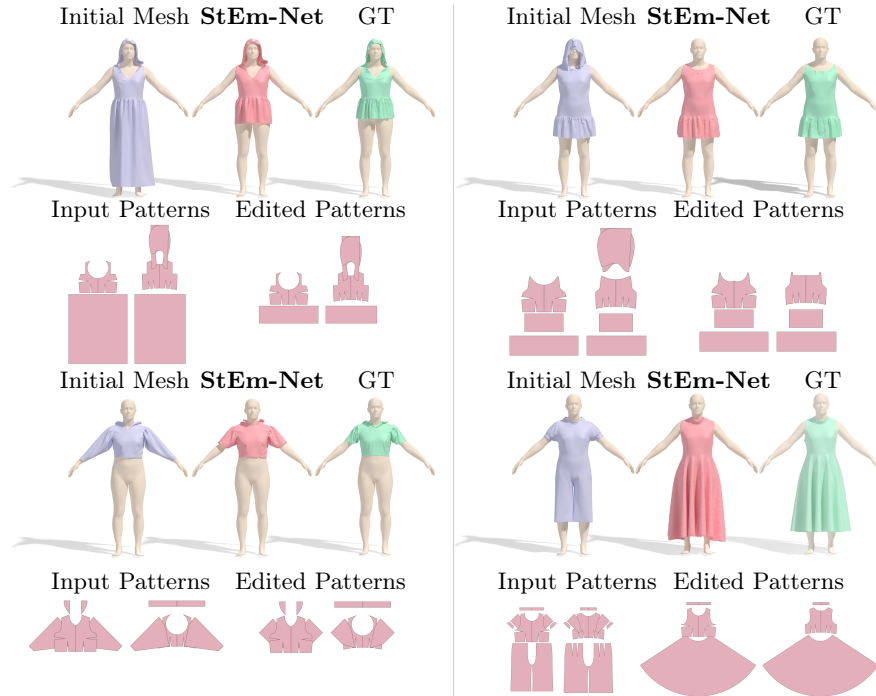


Fig. 8: Editing examples. For each example, we show on the left the initial sewing pattern with its 3D physical simulation (blue garment). On the right, the edited sewing pattern, with our network’s prediction (red) and the ground truth (green). We observe that our framework provides an accurate estimation of the garment’s appearance without requiring physical simulation, both for global behaviors (*e.g.*, length of the garment’s components) and for the addition or removal of garment parts (*e.g.*, replacing trousers with a skirt or removing sleeves and the hood).

	CD_{STP}	CD_{PTS}	CD_{bidir}
NeuralTailor	2.54	1.26	1.90
StEm-Net + Optim	0.96	0.67	0.81

Table 3: Mean Chamfer Distance for **Scan-To-Pattern**, **Pattern-To-Scan**, and Bidirectional Chamfer Distance (CD). We collected 543 CloSe scans and reposed them to the canonical pose.

5.5 Ablation on Input Normals

To evaluate our model’s robustness, we analyze the 3D reconstruction error under varying levels of Gaussian noise applied to the input BoxMesh. As illustrated in Fig. 10, we introduce noise to both the panel displacements (left) and the normal directions (right). In both scenarios, our method demonstrates significant resilience; the reconstruction error increases sublinearly relative to the noise level, indicating that the model remains stable and does not strictly overfit to pristine synthetic inputs.

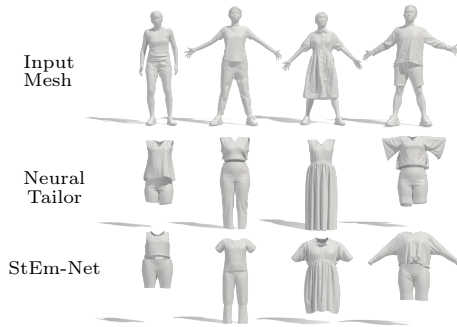


Fig. 9: Qualitative comparison on scans from 4DDress. StEm-Net reconstructs garments that adhere more closely to the input geometry than NeuralTailor.

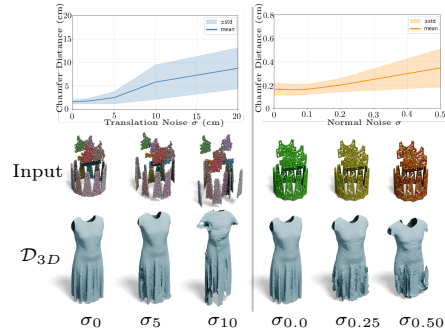


Fig. 10: We report the chamfer distance degradation with different Gaussian noise levels for BoxMesh translation (left) and normal orientation (right). Point color indicates the panel on the left, and the amount of per-point noise on the right.

6 Conclusions

In this work, we introduce StEm-Net, a novel framework that, for the first time, unifies 2D sewing patterns and 3D garments in a common latent space, here called *Stitched Embeddings*. This is achieved by representing sewing patterns as BoxMeshes, thereby making their connection with the 3D garment more direct and simpler to learn. Our method is fully differentiable, end-to-end trained, and does not require a physical simulator for test-time optimization. We showcase our advantages on two different datasets, improving the state of the art in sewing pattern reconstruction from 3D point clouds. Our framework supports test-time adaptation via backpropagation, making it highly computationally efficient. Within the same framework, we also enable direct propagation of sewing pattern edits to the 3D mesh, providing a practical, flexible tool for manufacturing. While our results are still not fully aligned with those from physical simulation, we view our work as an important milestone toward scalable digital garments, closing the gap between reconstruction and tailoring pipelines. Scaling to multi-layered attire, different materials, varying body models, and fine-grained nuances, such as pockets or ethnic garments, would be an interesting follow-up to our work. The principal obstacle in this direction is the scarcity of high-quality garment datasets.

Acknowledgements. We thank the whole RVH team for the support. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Andrea Sanchietti. This project was supported by Meta. Gerard Pons-Moll was supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ:01IS18039A, by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 409792180 (Emmy Noether Programme, project: Real Virtual Humans). GPM is a member of the Machine Learning Cluster of Excellence, EXC number 2064/1 - Project number 390727645, and is supported by the Carl Zeiss Foundation.

References

1. Antić, D., Tiwari, G., Ozcomlekci, B., Marin, R., Pons-Moll, G.: Close: A 3d clothing segmentation dataset and model. In: 2024 international conference on 3D vision (3DV). pp. 591–601. IEEE (2024)
2. Bang, S., Korosteleva, M., Lee, S.H.: Estimating garment patterns from static scan data. In: Computer Graphics Forum. vol. 40, pp. 273–287. Wiley Online Library (2021)
3. Berthouzoz, F., Garg, A., Kaufman, D.M., Grinspun, E., Agrawala, M.: Parsing sewing patterns into 3d garments. *Acm Transactions on Graphics (TOG)* **32**(4), 1–12 (2013)
4. Bhatnagar, B.L., Tiwari, G., Theobalt, C., Pons-Moll, G.: Multi-garment net: Learning to dress 3d people from images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
5. Bian, S., Xu, C., Xiu, Y., Grigorev, A., Liu, Z., Lu, C., Black, M.J., Feng, Y.: Chatgarment: Garment estimation, generation and editing via large language models (2025)
6. Can, S.E., Ackermann, J., Nakayama, K., Liu, R., Wu, T., Zheng, Y., Bertiche, H., Chai, M., Beeler, T., Wetzstein, G.: Image2garment: Simulation-ready garment generation from a single image. *arXiv preprint arXiv:2601.09658* (2026)
7. Chen, C.H., Su, J.W., Hu, M.C., Yao, C.Y., Chu, H.K.: Panelformer: Sewing pattern reconstruction from 2d garment images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 454–463 (January 2024)
8. Chen, X., Zhou, B., Lu, F., Wang, L., Bi, L., Tan, P.: Garment modeling with a depth camera. *ACM Transactions on Graphics (TOG)* **34**(6), 1–12 (2015)
9. Chibane, J., Mir, A., Pons-Moll, G.: Neural unsigned distance fields for implicit function learning. In: Advances in Neural Information Processing Systems (NeurIPS) (December 2020)
10. De Luigi, L., Li, R., Guillard, B., Salzmann, M., Fua, P.: Drapenet: Garment generation and self-supervised draping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1451–1460 (2023)
11. Deitke, M., Liu, R., Wallingford, M., Ngo, H., Michel, O., Kusupati, A., Fan, A., Laforte, C., Voleti, V., Gadre, S.Y., et al.: Objaverse-xl: A universe of 10m+ 3d objects. *Advances in Neural Information Processing Systems* **36**, 35799–35813 (2023)
12. Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., Vanderbilt, E., Schmidt, L., Ehsani, K., Kembhavi, A., Farhadi, A.: Objaverse: A universe of annotated 3d objects. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13142–13153 (2023)
13. Eskandar, D., Kabadayi, B., Tiwari, G., Pons-Moll, G.: Dama: Disentangled body-anchored gaussians for controllable multi-layered avatars. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 5799–5811 (June 2026)
14. Garavaso, D., Masi, F., Musoni, P., Castellani, U.: Point cloud segmentation for 3d clothed human layering. *Computers & Graphics* p. 104393 (2025)
15. Gong, J., Ji, S., Foo, L.G., Chen, K., Rahmani, H., Liu, J.: Laga: Layered 3d avatar generation and customization via gaussian splatting. *arXiv preprint arXiv:2405.12663* (2024)

16. Gu, X., Gortler, S.J., Hoppe, H.: Geometry images. In: Proceedings of the 29th annual conference on Computer graphics and interactive techniques. pp. 355–361 (2002)
17. Guo, J., Chen, J., Chen, W., Sun, Z., Li, L., Zhao, B., Zhu, L., Wang, X., Liu, Q.: Garmentx: Autoregressive parametric representations for high-fidelity 3d garment generation (2025)
18. Guo, M., Chiang, M.J.Y., Santesteban, I., Sarafianos, N., Chen, H.y., Halimi, O., Božič, A., Saito, S., Wu, J., Liu, C.K., Stuyck, T., Larionov, E.: Pgc: Physics-based gaussian cloth from a single pose. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 21215–21225 (June 2025)
19. He, K., Yao, K., Zhang, Q., Yu, J., Liu, L., Xu, L.: Dresscode: Autoregressively sewing and generating garments from text guidance. *ACM Transactions on Graphics (TOG)* **43**(4), 1–13 (2024)
20. He, Y., Xie, H., Miyata, K.: Sketch2cloth: Sketch-based 3d garment generation with unsigned distance fields. In: 2023 nicograph international (NicoInt). pp. 38–45. IEEE (2023)
21. Heming, Z., Yu, C., Hang, J., Weikai, C., Dong, D., Zhangye, W., Shuguang, C., Xiaoguang, H.: Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images. In: *Computer Vision – ECCV 2020*. pp. 512–530. Springer International Publishing (2020)
22. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* **9**(8), 1735–1780 (1997)
23. Howell, T.A., Le Cleac’h, S., Kolter, J.Z., Schwager, M., Manchester, Z.: Dojo: A differentiable simulator for robotics. *arXiv preprint arXiv:2203.00806* **9**(2), 4 (2022)
24. Jin, L., Jin, Z., Ye, Z., Pang, H., Han, X., Zheng, Y., Li, H.: Inversedraping: Recovering sewing patterns from 3d garment surfaces via boxmesh bridging. *IEEE Transactions on Visualization and Computer Graphics* **32**(7), 6350–6362 (2026)
25. Jung, M., Lee, D., Lee, I.K.: Clothingtwin: Reconstructing inner and outer layers of clothing using 3d gaussian splatting. In: *Computer Graphics Forum*. vol. 44, p. e70240. Wiley Online Library (2025)
26. Korosteleva, M., Kesdogan, T.L., Kemper, F., Wenninger, S., Koller, J., Zhang, Y., Botsch, M., Sorkine-Hornung, O.: GarmentCodeData: A dataset of 3D made-to-measure garments with sewing patterns. In: *Computer Vision – ECCV 2024* (2024)
27. Korosteleva, M., Lee, S.H.: Generating datasets of 3d garments with sewing patterns. In: Vanschoren, J., Yeung, S. (eds.) *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*. vol. 1 (2021)
28. Korosteleva, M., Lee, S.H.: Neuraltailor: Reconstructing sewing pattern structures from 3d point clouds of garments. *ACM Trans. Graph.* **41**(4) (2022)
29. Korosteleva, M., Sorkine-Hornung, O.: GarmentCode: Programming parametric sewing patterns. *ACM Transaction on Graphics* **42**(6) (2023), sIGGRAPH ASIA 2023 issue
30. Li, M., Shan, H., Zheng, K., Shen, C., Liu, S., Fu, Y., Chen, Z., Huang, X.: Reweaver: Towards simulation-ready and topology-accurate garment reconstruction. *arXiv preprint arXiv:2601.16672* (2026)
31. Li, S., Liu, R., Liu, C., Wang, Z., He, G., Li, Y.L., Jin, X., Wang, H.: Garmagenet: A multimodal generative framework for sewing pattern design and generic garment modeling. *ACM Transactions on Graphics (TOG)* **44**(6), 1–23 (2025)

32. Li, X., Yao, Q., Wang, Y.: Garmentdiffusion: 3d garment sewing pattern generation with multimodal diffusion transformers. In: Kwok, J. (ed.) *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*. pp. 1458–1466. International Joint Conferences on Artificial Intelligence Organization (8 2025), main Track
33. Li, Y., Chen, H.y., Larionov, E., Sarafianos, N., Matusik, W., Stuyck, T.: DiffAvatar: Simulation-ready garment optimization with differentiable simulation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2024)
34. Liu, L., Xu, X., Lin, Z., Liang, J., Yan, S.: Towards garment sewing pattern reconstruction from a single image. *ACM Transactions on Graphics (SIGGRAPH Asia)* (2023)
35. Liu, S., Cheng, Y., Chen, Z., Ren, X., Zhu, W., Li, L., Bi, M., Yang, X., Yan, Y.: Multimodal latent diffusion model for complex sewing pattern generation. *International Conference on Computer Vision (ICCV)* (2025)
36. Liu, X., Li, J., Lu, G.: Reconstructing complex shaped clothing from a single image with feature stable unsigned distance fields. *IEEE Transactions on Visualization and Computer Graphics* **31**(4), 2142–2154 (2024)
37. Liu, Y., Tang, J., Zheng, C., Zhu, J., Wang, C., Huang, D.: Clothedreamer: Text-guided garment generation with 3d gaussians: Y. liu et al. *Applied Intelligence* **55**(10), 767 (2025)
38. Liu, Z., Feng, Y., Xiu, Y., Liu, W., Paull, L., Black, M.J., Schölkopf, B.: Ghost on the shell: An expressive representation of general 3d shapes. In: *ICLR* (2024)
39. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* **34**(6), 248:1–248:16 (Oct 2015)
40. Luo, J., Qu, H., Zhao, Y., Zhang, J., Yang, Y.: Deep learning for 3d fashion design: A survey from a sewing pattern-driven perspective. *IEEE Transactions on Circuits and Systems for Video Technology* (2025)
41. Macklin, M.: Warp: A high-performance python framework for gpu simulation and graphics. In: *NVIDIA GPU Technology Conference (GTC)*. vol. 3 (2022)
42. Mir, A., Guler, R.A., Tang, X., Wonka, P., Pons-Moll, G.: Ahoy! animatable humans under occlusion from youtube videos with gaussian splatting and video diffusion priors. In: *European Conference on Computer Vision (ECCV)* (2026)
43. Musoni, P., Melzi, S., Castellani, U.: Gim3d: A 3d dataset for garment segmentation. In: *Smart Tools and Applications in Graphics - Eurographics Italian Chapter Conference*. The Eurographics Association (2022)
44. Musoni, P., Melzi, S., Castellani, U.: Gim3d plus: A labeled 3d dataset to design data-driven solutions for dressed humans. *Graphical models* **129**, 101187 (2023)
45. Musoni, P., Melzi, S., Castellani, U.: Capturing and modeling real cloth deformations for virtual garment design. In: *European Conference on Computer Vision*. pp. 320–336. Springer (2024)
46. Nakayama, K., Ackermann, J., Kesdogan, T.L., Zheng, Y., Korosteleva, M., Sorkine-Hornung, O., Guibas, L.J., Yang, G., Wetzstein, G.: Aipparel: A multi-modal foundation model for digital garments. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 8138–8149 (2025)
47. Nakayama, K., Shen, I.C., Liu, R., Wang, Y., Wetzstein, G., Igarashi, T.: Garment particles: A 2d–3d symmetric garment representation for generation and editing. In: *SIGGRAPH Conference Papers '26* (2026)

48. Pietroni, N., Dumery, C., Falque, R., Liu, M., Vidal-Calleja, T.A., Sorkine-Hornung, O.: Computational pattern making from 3d garment models. *ACM Trans. Graph.* **41**(4), 157–1 (2022)
49. Pons-Moll, G., Pujades, S., Hu, S., Black, M.J.: Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics (ToG)* **36**(4), 1–15 (2017)
50. Rong, B., Grigorev, A., Wang, W., Black, M.J., Thomaszewski, B., Tsalicoglou, C., Hilliges, O.: Gaussian garments: Reconstructing simulation-ready clothing with photorealistic appearance from multi-view video. In: 2025 International Conference on 3D Vision (3DV). pp. 1054–1063. IEEE (2025)
51. Shen, Y., Liang, J., Lin, M.C.: Gan-based garment generation using sewing pattern images. In: European Conference on Computer Vision. pp. 225–247. Springer (2020)
52. Tatsukawa, Y., Qi, A., Shen, I.C., Igarashi, T.: Garmentimage: Raster encoding of garment sewing patterns with diverse topologies. In: ACM SIGGRAPH 2025 Conference Proceedings (2025)
53. Team, T.H.: Hunyuan3d 2.1: From images to high-fidelity 3d assets with production-ready pbr material (2025)
54. Tiwari, G., Bhatnagar, B.L., Tung, T., Pons-Moll, G.: Sizer: A dataset and model for parsing 3d clothing and learning size sensitive 3d clothing. In: European Conference on Computer Vision. pp. 1–18. Springer (2020)
55. Tiwari, G., Sarafianos, N., Tung, T., Pons-Moll, G.: Neural-gif: Neural generalized implicit functions for animating people in clothing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11708–11718 (2021)
56. Vuran, O., Ho, H.I.: Remu: Reconstructing multi-layer 3d clothed human from images. In: British Machine Vision Conference (BMVC) (2025)
57. Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., Novotny, D.: Vggt: Visual geometry grounded transformer. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5294–5306 (2025)
58. Wang, K., Wang, C., Yang, J., Zhang, G.: Clocap-gs: clothed human performance capture with 3d gaussian splatting. *IEEE Transactions on Image Processing* (2025)
59. Wang, T.Y., Ceylan, D., Popovic, J., Mitra, N.J.: Learning a shared shape space for multimodal garment design. *arXiv preprint arXiv:1806.11335* (2018)
60. Wang, W., Ho, H.I., Guo, C., Rong, B., Grigorev, A., Song, J., Zarate, J.J., Hilliges, O.: 4d-dress: A 4d dataset of real-world human clothing with semantic annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 550–560 (2024)
61. Wang, Y., Zhang, C., Frazão, G., Yang, J., Ichim, A.E., Beeler, T., De la Torre, F.: Garmentcrafter: Progressive novel view synthesis for single-view 3d garment reconstruction and editing. *arXiv preprint arXiv:2503.08678* (2025)
62. Wen, M., Wu, S., Wang, K., Liang, D.: Intergsedit: Interactive 3d gaussian splatting editing with 3d geometry-consistent attention prior. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 26136–26145 (2025)
63. Wolf, Y., Bracha, A., Kimmel, R.: Gs2mesh: Surface reconstruction from gaussian splatting via novel stereo views. In: European Conference on Computer Vision. pp. 207–224. Springer (2024)
64. Wolff, K., Herholz, P., Sorkine-Hornung, O.: Reflection symmetry in textured sewing patterns. In: VMV. pp. 11–18 (2019)
65. Wolff, K., Herholz, P., Ziegler, V., Link, F., Brügel, N., Sorkine-Hornung, O.: Designing personalized garments with body movement. In: Computer Graphics Forum. vol. 42, pp. 180–194. Wiley Online Library (2023)

66. Wu, Q., Zheng, J., Cai, J.: Surface reconstruction from 3d gaussian splatting via local structural hints. In: European Conference on Computer Vision. pp. 441–458. Springer (2024)
67. Xiang, D., Prada, F., Wu, C., Hodgins, J.: Monoclothcap: Towards temporally coherent clothing capture from monocular rgb video. In: 2020 International Conference on 3D Vision (3DV). pp. 322–332. IEEE (2020)
68. Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., Chen, D., Tong, X., Yang, J.: Structured 3d latents for scalable and versatile 3d generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 21469–21480 (2025)
69. Xue, Y., Xie, X., Marin, R., Pons-Moll, G.: Human-3diffusion: realistic avatar creation via explicit 3d consistent diffusion models. *Advances in Neural Information Processing Systems* **37**, 99601–99645 (2024)
70. Xue, Y., Xie, X., Marin, R., Pons-Moll, G.: Gen-3diffusion: Realistic image-to-3d generation via 2d & 3d diffusion synergy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025)
71. Xueting, L., Yuan, Y., De Mello, S., Daviet, G., Leaf, J., Macklin, M., Kautz, J., Iqbal, U.: Simavatar: Simulation-ready avatars with layered hair and clothing. *Arxiv* (2024)
72. Yang, S., Ambert, T., Pan, Z., Wang, K., Yu, L., Berg, T., Lin, M.C.: Detailed garment recovery from a single-view image. *arXiv preprint arXiv:1608.01250* (2016)
73. Yu, B., Cordier, F., Seo, H.: Inverse garment and pattern modeling with a differentiable simulator. *Computer Graphics Forum* (2024)
74. Yu, Z., Dou, Z., Long, X., Lin, C., Li, Z., Liu, Y., Müller, N., Komura, T., Habermann, M., Theobalt, C., et al.: Surf-d: High-quality surface generation for arbitrary topologies using diffusion models. *arXiv preprint arXiv:2311.17050* (2023)
75. Zhang, C., Pujades, S., Black, M.J., Pons-Moll, G.: Detailed, accurate, human shape estimation from clothed 3d scan sequences. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4191–4200 (2017)
76. Zhang, J., Huang, J., Jin, S., Lu, S.: Vision-language models for vision tasks: A survey. *IEEE transactions on pattern analysis and machine intelligence* **46**(8), 5625–5644 (2024)
77. Zhang, X., Liu, Z., Zhang, Y., Ge, X., He, D., Xu, T., Wang, Y., Lin, Z., Yan, S., Zhang, J.: Mega: Memory-efficient 4d gaussian splatting for dynamic scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 27828–27838 (2025)
78. Zhou, F., Liu, R., Liu, C., He, G., Li, Y.L., Jin, X., Wang, H.: Design2garmentcode: Turning design concepts to tangible garments through program synthesis. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 23712–23722 (2025)
79. Zou, X., Han, X., Wong, W.: Cloth4d: A dataset for clothed human reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12847–12857 (2023)