Image2Garment: Simulation-ready Garment Generation from a Single Image

Selim Emir Can, Jan Ackermann, Kiyohiro Nakayama, Ruofan Liu, Tong Wu, Yang Zheng, Hugo Bertiche, Menglei Chai, Thabo Beeler, Gordon Wetzstein

Stanford University | Google

Computational Imaging (EE367) — Project Proposal

1. Motivation

Generating physically accurate, simulation-ready garments from casual, in-the-wild images is one of the outstanding challenges at the intersection of computational imaging, computer vision, and physically based simulation. Applications in virtual reality, gaming, and fashion design all demand garment representations that are not only geometrically plausible but also endowed with the material properties — bending stiffness, stretch, shear, damping — necessary for realistic cloth dynamics.

Existing approaches face a fundamental data bottleneck: no large-scale dataset pairs real-world garment images with the physics parameters required by cloth simulators, and gathering such data requires prohibitive manual measurement. Methods that do recover physical parameters rely on multi-view video capture and expensive iterative differentiable simulation, which are impractical outside controlled studio environments. Conversely, the recent wave of single-image garment generation methods reconstruct compelling 3D shapes but produce no material information, forcing users to guess physics parameters manually.

Our key insight is to reformulate this ill-posed inverse problem through a semantically grounded latent decomposition. Although direct image-to-physics supervision is unavailable, image-to-material information is abundant: large online retail catalogs provide reliable material-composition labels (e.g., cotton 95%, elastane 5%) together with fabric-family and weave-structure metadata. These mid-level descriptors occupy a structured, low-dimensional space with a far more predictable relationship to simulator parameters than raw pixel values do, making the mapping substantially easier to learn from a modest collection of material–physics measurements.

2. Related Work

Single-image garment generation. Recent generative methods, driven by fine-tuned vision–language models (VLMs), have substantially improved single-image garment reconstruction. ChatGarment [1] generates sewing patterns from a single image using large language models, enabling downstream simulation-ready draping on SMPL bodies. AIpparel [2], a multimodal foundation model from our own lab, recovers detailed garment meshes and sewing patterns from in-the-wild images. While these methods produce compelling geometry, they predict no material properties, requiring users to specify physics parameters manually for every fabric panel. GarmentRecovery [3] reconstructs garment meshes using shape and deformation priors but similarly produces no physics parameters compatible with standard cloth simulators.

Physics parameter estimation. Methods that do recover physics parameters from visual data require strong supervision from multi-view or video setups. Dress-1-to-3 [4] estimates sewing patterns and physics parameters from a single image but relies on hours of iterative differentiable simulation optimization per garment and outputs a proprietary parameter format not compatible with commercial simulators. PhysAvatar [5] and GaussianGarments [6] achieve high-fidelity results by optimizing garment geometry and physics jointly from multi-view video, but their reliance on controlled capture makes them inaccessible for casual users. Our work avoids iterative optimization entirely, producing simulation-ready outputs in a single feedforward pass.

Fabric mechanics estimation. Wang et al. [7] fit elastic models from physical measurements of real fabric samples, and Dominguez-Elvira et al. [8] demonstrate a practical method to estimate fabric mechanics from material metadata. Crucially, neither of these works connect fabric mechanics estimation to in-the-wild visual observation, a gap that our two-stage framework is specifically designed to bridge.

References

- [1] Bian et al. ChatGarment: Garment Estimation, Generation and Editing via Large Language Models. CVPR 2025.
- [2] Nakayama et al. AIpparel: A Multimodal Foundation Model for Digital Garments. CVPR 2025.
- [3] Li et al. Garment Recovery with Shape and Deformation Priors. CVPR 2024.
- [4] Li et al. Dress-1-to-3: Single Image to Simulation-ready 3D Outfit with Diffusion Prior and Differentiable Physics. ACM TOG 2025.
- [5] Zheng et al. PhysAvatar: Learning the Physics of Dressed 3D Avatars from Visual Observations. ECCV 2024.
- [6] Rong et al. Gaussian Garments: Reconstructing Simulation-ready Clothing with Photorealistic Appearance from Multi-view Video. 3DV 2025.
- [7] Wang et al. Data-Driven Elastic Models for Cloth: Modeling and Measurement. ACM TOG 2011.
- [8] Dominguez-Elvira et al. Practical Method to Estimate Fabric Mechanics from Metadata. CGF 2024.

3. Project Overview

We propose Image2Garment, a fully feed-forward, optimization-free framework that generates simulation-ready garments from a single RGB photograph. Our pipeline decomposes the otherwise ill-posed image-to-physics problem into three learnable stages:

- Stage 1 — Garment Geometry. We use ChatGarment as a frozen backbone to predict a sewing-pattern representation from the input image, which is then draped on a canonical SMPL body mesh using a standard cloth simulator.
- Stage 2 — Fabric Attribute Estimation. We fine-tune Qwen-2.5VL, a vision–language model, on our Fabric Attributes from Garment Tags (FTAG) dataset — 16,026 images paired with vendor-provided material composition, fabric family, and structure type labels. The fine-tuned VLM outputs a structured JSON containing per-fiber percentages, fabric family, and structure type. Areal density and thickness are subsequently estimated via hierarchical retrieval from our Tag-to-Physics (T2P) dataset.
- Stage 3 — Physics Parameter Prediction. A collection of lightweight Random Forest Regressors maps the predicted fabric attributes to the full set of simulator-compatible physics parameters (bending, shear, stretch, and buckling stiffness; damping; friction), following the practical recipe of Dominguez-Elvira et al. [8]. The predicted parameters are directly importable into commercial garment design software such as Marvelous Designer / CLO3D and Browzwear.

Datasets

- FTAG (Fabric Attributes from Garment Tags): 16,026 garment images annotated with material composition, fabric family (38 classes), and structure type (knit / woven / other), curated from public online retail sources.
- T2P (Tag-to-Physics): 1,254 digitized fabrics linking vendor-provided fabric attributes to CLO3D-compatible physics parameters measured with the CLO Fabric Kit.

Final Goals

- Demonstrate that a semantically grounded latent decomposition (image → fabric attributes → physics parameters) transforms an ill-posed inverse problem into two well-posed, data-efficient supervised tasks.
- Outperform state-of-the-art single-image garment methods (ChatGarment, AIpparel, GarmentRecovery, Dress-1-to-3) on both geometry quality (Chamfer Distance, IoU) and dynamic draping fidelity (PSNR, SSIM, LPIPS) across multiple animation sequences.

- Achieve superior fabric attribute estimation compared to strong VLM baselines (GPT-5, zero-shot and few-shot), demonstrating the value of domain-specific fine-tuning.
- Produce a fully open, feedforward pipeline that takes seconds per garment versus hours for optimization-based competitors.

4. Milestones & Timeline

Phase	Target Date	Deliverable / Milestone
Geometry Baseline	Week 1	Integrate ChatGarment as frozen sewing-pattern backbone; establish evaluation pipeline using CLO3D simulation and Blender rendering; reproduce baseline geometry metrics (CD, IoU).
Fabric Attribute Estimator	Week 1	Fine-tune Qwen-2.5VL with LoRA (rank 64) on FTAG using inverse-frequency token weighting; implement hierarchical density–thickness retrieval; validate on FTAG test split (1,231 samples).
Physics Predictor	Week 2	Train Random Forest Regressors for bending, shear, stretch, and buckling stiffness on T2P; tune hyperparameters via 50-iteration randomized search with 5-fold cross-validation; report NMAE.
Full Pipeline Integration	Week 2	Connect all three stages into a single feedforward pipeline; create synthetic evaluation dataset (4 outfits × 4 Mixamo animations, rendered in Blender); run end-to-end evaluation.
Comparative Evaluation	Week 2	Evaluate against ChatGarment*, AIpparel*, GarmentRecovery*, and Dress-1-to-3 on both the synthetic dataset and the 4D-Dress benchmark; report Table 1 and Table 3 results.
Ablations & Write-up	Week 3	Conduct ablation on physics parameters and fabric attribute estimation; Adress CVPR26 revisions
Code and Data Release	Week 2-3	Create a GitHub code repository and release collected data as a Json of links to garments upon acceptance to CVPR26