Image2Garment: Simulation-ready Garment Generation from a Single Image

Selim Emir Can

Stanford University, Stanford CA 94305, USA
emircan@stanford.edu

Abstract. Estimating physically accurate, simulation-ready garments from a single image is challenging due to the absence of image-to-physics datasets and the inherently ill-posed nature of the problem. Prior methods either require multi-view capture and expensive differentiable simulation, or recover garment geometry without the material properties required for realistic simulation. We propose a feed-forward framework that predicts simulation-ready garments from a single image. Our method first fine-tunes a vision–language model to estimate material composition and fabric attributes from real images. We then train a lightweight predictor that maps these attributes to physical fabric parameters using a dataset of material–physics measurements. To support this pipeline, we introduce three new datasets for training and evaluation. Experiments show that our approach achieves superior accuracy in material composition estimation and fabric attribute prediction. By estimating physical parameters from these predictions, our method produces higher-fidelity simulations than existing image-to-garment approaches. We will release all datasets and trained models to facilitate reproducibility.

1 Introduction

Generating simulation-ready garments from visual observations is increasingly important for applications in virtual reality, gaming, and fashion design. Although recent methods can recover physical fabric properties, they typically rely on multi-view capture setups or manual material measurement, both of which are labor-intensive and impractical outside controlled environments. Single-image inference would be a far more accessible alternative, but estimating a garment's physical properties (e.g., stretch and bend stiffness, density, damping) from an in-the-wild image remains challenging due to limited viewpoints, ambiguous drape cues, and the lack of direct supervision.

Recent progress in single-image garment generation [9, 25, 27, 29, 38] has substantially improved garment geometry and appearance prediction. However, these methods largely neglect the physical parameters necessary for faithful simulation (see Fig. 1). Conversely, methods that optimize physical parameters via differentiable simulation [15, 25, 26, 30, 37] require multi-view inputs, involve slow iterative optimization, and are typically restricted to simple garment categories. Obtaining simulation-ready garments from a single image thus remains unsolved.

Our key insight is to reformulate this inverse problem through a semantically grounded latent decomposition. Although direct image-to-physics supervision is unavailable, image-to-material information is abundant: online clothing catalogs provide reliable material-composition labels (e.g., cotton, silk, polyester blends) and complementary attributes such as fabric family and weave structure. These descriptors occupy a low-dimensional, structured space with a far more predictable relationship to physical parameters than raw images do. The material-to-physics mapping is therefore substantially easier to learn, requiring only a modest dataset of material–physics measurements.

We introduce a two-stage factorization: (i) train a model to infer interpretable material descriptors from a single input image; (ii) using independently gathered material–physics annotations, train a mapper converting these descriptors to full simulator parameters. This latent-variable formulation regularizes the learning task, dramatically reduces data requirements, and resolves the ill-posedness that makes direct image-to-physics prediction impractical.

In summary, our contributions are:

– **Three new datasets**: the Fabric Attributes from Garment Tags (FTAG) dataset (16,026 images annotated with material composition, fabric family, and structure type), the Tag-to-Physics (T2P) dataset

Fig. 1: Impact of fabric parameters on simulation. Four different fabrics starting from the same initial condition produce drastically different dynamics, highlighting the critical importance of accurate physics parameter estimation.

(1,354 fabrics linking attributes to simulator-compatible physics parameters), and the Synthetic Motion Sequences (SMS) dataset for evaluating garment dynamics.
- **A feed-forward pipeline** for single-image to simulation-ready garment generation, jointly predicting geometry and fabric descriptors, then mapping them to physically realistic material parameters.
- **Extensive experiments** demonstrating state-of-the-art fabric attribute estimation and superior garment simulation quality compared to baselines that use randomly sampled physics parameters.

2 Related Work

2.1 Garment Reconstruction and Generation.

Many works reconstruct garments from multi-view images and videos [2, 8, 14, 18, 20], but do not model material composition and therefore cannot directly drive physics simulation. Other approaches [15, 30, 37] use inverse rendering and physics to jointly optimize geometry and physical parameters, but require studio-quality video, making them impractical for casual users.

Generative models have enabled single-image garment generation [3,31,34,35], producing 3D meshes that require post-processing to separate garment from human. Others model garment and human separately [12, 24,28] but still lack simulation-compatible physics parameters. Another line of work directly generates sewing patterns from images [9, 16, 23, 27, 29, 32]; while more simulation-ready, users still must specify physical parameters manually.

The work closest to ours is Dress 1-to-3 [25], which estimates both sewing pattern and physical parameters from a single image via inverse optimization. However, it requires hours of optimization per garment and is restricted to differentiable simulators.

2.2 Estimating Physics Parameters from Visual Data.

Estimating physics from visual observations is an increasingly important topic [6,22,36]. Several works obtain garment physics parameters from video [7, 13, 25, 30, 37], but all require multi-view or video setups and most are incompatible with established cloth simulators. High-quality single-image estimation of garment physical parameters remains largely unexplored.

3 Method

Overview. Our goal is to recover a simulation-ready garment $G = (S, \boldsymbol{\theta})$ from a single RGB image I, where S is the garment geometry and $\boldsymbol{\theta}$ are the physical simulator parameters. We seek:

$$\hat{G} = \arg\max_{S,\boldsymbol{\theta}} p(S, \boldsymbol{\theta} \mid I). \tag{1}$$

Our key insight is to introduce latent fabric attributes Z (material composition C, structure type s, fabric family f) and decompose:

$$p(S, \boldsymbol{\theta} \mid I) = p(S \mid I) \sum_{z} p(\boldsymbol{\theta} \mid z)\, p(Z = z \mid I). \tag{2}$$

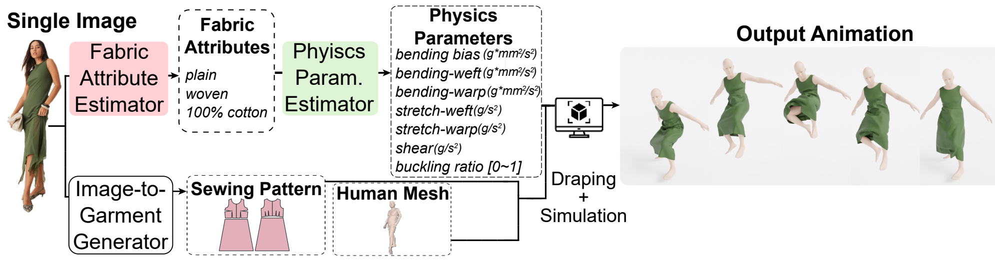

Fig. 2: Image2Garment pipeline. From a single image: (1) generate garment sewing pattern via ChatGarment [9], (2) predict fabric attributes (material, family, structure), (3) map to physical parameters for realistic simulation.

This concentrates uncertainty into $p(Z \mid I)$—where rich supervision exists—and turns physics estimation into a well-posed supervised mapping. We implement three stages:

$$\hat{S} = \arg\max_{S} p(S \mid I), \quad \hat{Z} = \arg\max_{z} p(z \mid I), \quad \hat{\boldsymbol{\theta}} = f_\phi(\hat{Z}). \tag{3}$$

3.1 Datasets

FTAG: Fabric Attributes from Garment Tags. We curate 16,026 images of retail garments paired with material composition (fiber type and percentages), fabric family (e.g., denim, jersey, chiffon), and structure type (knit/woven/other). Images were collected from public online sources and processed through extensive cleaning and normalization, yielding 9,843/1,231/1,231 train/val/test samples.

Tag-to-Physics (T2P). We curate 1,359 fabrics paired with CLO3D physics parameters: bending, shear, stretch, and buckling stiffness (warp/weft/bias directions), buckling ratio, areal density, thickness, friction, and damping. The dataset is split using 5-fold stratified cross-validation (≈864/160/335 train/val/test per fold).

Synthetic Motion Sequence (SMS). The SMS dataset contains five motion sequences (jumping jack, spinning, jab cross, joyful jump, hit reaction) with garments from GarmentCodeData [19] animated via Mixamo [1] and simulated in CLO3D with realistic materials (polyester, cotton, wool). The first frame is used as input and metrics are computed over all subsequent frames. We additionally include three in-the-wild videos (cultural dance, movie clip, contemporary dance) for qualitative evaluation only, using an off-the-shelf SMPL body fitting method to recover motion sequences.

4D-Dress dataset. In addition to the SMS dataset, we also evaluate garment reconstruction and dynamic draping quality on the 4D-Dress dataset [33]. The dataset contains 50 multi-view video sequences of clothed subjects from which we select 5 sequences for evaluation. Following [4], this subset includes the ChatGarment split of 4D-Dress [9] which consists of 4 sequences.

3.2 Simulation-ready Garment Pipeline

Our pipeline has three stages. First, we reconstruct garment geometry by estimating its sewing pattern using ChatGarment [9] and draping it with a standard cloth simulator [21]. Next, we predict fabric attributes (material composition, fabric family, structure type) from the input image. Finally, we map these descriptors to simulator-compatible physical parameters. See Fig. 2 for an overview.

Fabric Attribute Prediction. We formulate fabric attribute estimation as an image captioning task, prompting a VLM to output a structured JSON with per-fiber percentages, fabric family, and structure type. We fine-tune Qwen-2.5VL [5] with LoRA [17], using inverse class-frequency weighted cross-entropy to counter material imbalance in the FTAG dataset:

$$\mathcal{L}_{\mathrm{VLM}} = -\sum_{i} w_{t_i} \log p_\theta(t_i \mid \mathcal{I}, \mathbf{t}_{<i}).$$

Table 1: Quantitative comparison on SMS dataset (averaged over sequences). Best in **bold**, second underlined. Arrows indicate optimization direction. CD is multiplied by 10^4.

Method	CD↓	IoU↑	PSNR↑	SSIM↑	LPIPS↓
GarmentRecovery*	429.0	7.0	16.56	0.899	0.168
AIpparel*	126.0	17.5	21.79	0.946	0.075
ChatGarment*	189.0	19.0	21.75	0.940	0.083
Image2Garment (ours)	**90.8**	**21.5**	**22.50**	**0.946**	**0.072**

	Lower		Upper	
Method	CD ↓	IoU ↑	CD ↓	IoU ↑
Dress 1-to-3 [25]	34.1	44.8	40.5	40.0
ChatGarment [9]	28.9	48.4	28.1	46.6
AIpparel [29]	380	28.0	380	28.0
Ours	**27.8**	**48.7**	**27.4**	**47.4**

Table 2: Clothing reconstruction benchmark. We report the average Chamfer Distance (CD ↓) and Intersection over Union (IoU ↑) between the ground-truth garment meshes and the reconstructed clothing averaged over time for the lower and upper garment categories in 4D-Dress. Our method achieves the best score in both metrics, showcasing the importance of accurate physical parameters for garment dynamics. Reported CD is multiplied by 1e4. Dress123 failed on one sample, and AIpparel failed on two samples. AIpparel generates a single garment per image; we evaluate once against the union of Upper+Lower meshes and report it in both places.

Physics Parameter Prediction. The joint categorical fabric attributes explain on average $\eta^2 = 0.67$ of variance in physics parameters, confirming the mapping lies in a relatively low-dimensional, well-conditioned space. We predict physics parameters $\mathcal{P}$ from fabric attributes (C, f, s) using a joint Random Forest Regressor (RFR) [10]. The feature vector concatenates normalized fiber percentages and one-hot encodings of fabric family and structure type. Hyperparameters are selected via 50-iteration randomized search with 5-fold stratified cross-validation, minimizing MAE. The best model achieves 0.08 NMAE and 0.60 Spearman correlation on 335 test samples.

4 Analysis, Evaluation & Comparison

Setup. We fine-tune Qwen-2.5VL [5] on FTAG (9,843/1,231/1,231 split) with LoRA (rank 64) for 3000 steps using AdamW (lr 1×10^{-5}, cosine schedule). All simulations use Marvelous Designer [11] at 24 fps on SMPL bodies with identical settings across methods.

Datasets. **SMS**: Five synthetic motion sequences with garments from GarmentCodeData [19], animated with Mixamo [1], simulated in CLO3D, and rendered in Blender. **4D-Dress**: 12 single-layer real-world sequences from Wang et al. [33]. The first frame is used as input in both cases.

Metrics. Chamfer Distance (CD↓, $\times10^4$) and IoU (↑) for 3D shape; PSNR, SSIM, LPIPS for image quality.

Baselines. ChatGarment [9], AIpparel [29], GarmentRecovery [24], and Dress-1-to-3 [25]. None produce simulator-compatible physics, so we augment them with randomly sampled materials (*) to isolate the importance of accurate physics prediction.

Quantitative Comparison. Table 1 shows average SMS results. Our method achieves the best performance across all 3D metrics. Table 2 shows 4D-Dress results where our method achieves the best CD and IoU for both upper and lower garments, confirming that accurate physics prediction leads to better-matching cloth dynamics.

Table 3: Performance of fabric attribute estimation baselines. ChatGPT (zero-shot) and ChatGPT (few-shot) results are compared with our finetuned Qwen2.5-VL model. Categorical fields are evaluated using Accuracy and F1 Score, while continuous-valued fields are evaluated using MAE and NMAE (%). All fields are estimated on a stratified test split of our material composition estimation benchmark (FTAG).

Attribute Field	ChatGPT (zero-shot)	ChatGPT (few-shot)	Ours
Categorical Fields (Accuracy % / F1-score) ↑			
Fabric Family	0.58 / 0.42	<u>0.61</u> / <u>0.43</u>	**0.75** / **0.72**
Structure Type	0.74 / 0.68	<u>0.75</u> / <u>0.69</u>	**0.86** / **0.85**
Material Type	0.65 / <u>0.70</u>	<u>0.66</u> / <u>0.70</u>	**0.71** / **0.75**
Continuous Fields (MAE % / NMAE) ↓			
Material Percentage	23.3 / 0.45	<u>22.4</u> / <u>0.43</u>	**19.3** / **0.40**

Table 4: Physics parameter prediction ablation (averaged over sequences). Using ground-truth garment geometry with our predicted vs. randomly sampled physics. Best in **bold**. CD$\times 10^4$.

Method	CD↓	IoU↑	PSNR↑	SSIM↑	LPIPS↓
Random Parameters	49.41	31.80	25.56	0.957	0.041
Image2Garment (ours)	**10.96**	**51.68**	**29.98**	**0.972**	**0.018**

Table 2 shows that our method achieves the best CD and IoU across both upper and lower garments on 4D-Dress, outperforming the closest baseline (ChatGarment*) by a consistent margin. Notably, Dress 1-to-3 [25] — despite requiring hours of per-garment optimization — underperforms our feed-forward approach, confirming that accurate physics prediction is more effective than slow iterative fitting.

Fabric Attribute Estimation. Table 3 reports performance on 1,231 test samples. Our fine-tuned Qwen2.5-VL outperforms zero-shot and few-shot ChatGPT across all fields, with gains of up to 30% in material type accuracy and 20% MAE reduction in percentage estimation.

Ablations. Table 4 ablates physics prediction by fixing ground-truth garment geometry and varying only physics parameters. Our predicted parameters yield substantial improvements over randomly sampled materials (e.g., 5× lower CD, 1.6× higher IoU on average), confirming that accurate physics estimation is critical for realistic garment dynamics.

Table 5: Ablation of model architectures for physics parameter prediction. Comparison of physics parameter estimation methods on our fabric dataset (860 train / 160 val / 335 test samples).

Method	NMAE ↓	Spearman ↑
Random Forest Regressor	**0.08**	**0.60**
MLP (5-layer)	<u>0.10</u>	0.46
SVR (RBF kernel)	0.11	<u>0.58</u>

Physics Parameter Estimation. Random Forest achieves the best performance for mapping fabric attributes to physics parameters, outperforming SVR and MLP baselines. We attribute this to the heterogeneous, partially discrete nature of the attribute-to-physics mapping, where tree-based recursive partitioning naturally captures regime changes across fabric type combinations — something SVR's smooth kernel interpolation and MLP's high-capacity fitting both handle less reliably at this data scale. Results are shown in Table 5.

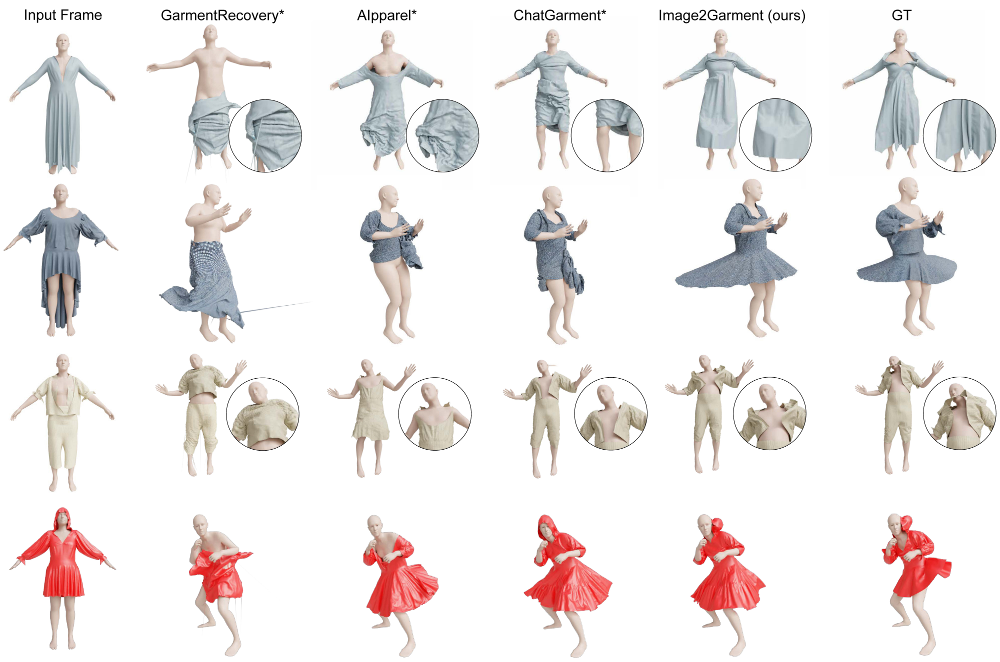

Fig. 3: Qualitative comparison. Our approach reconstructs garments that more faithfully match ground-truth geometry and cloth behavior. Competing methods produce artifacts, incorrect topology, or unrealistic draping. Zoomed-in regions highlight improvements in folds and garment structure.

5 Results

Qualitative Results. Figure 3 shows qualitative comparisons on the SMS dataset. Specifically, GarmentRecovery* produces garments with collapsing geometry. AIpparel* either produces incorrect garment geometry or incorrect wrinkles or unwanted folding during simulation. ChatGarment* produces unrealistic wrinkles and unwanted garment motion due to inaccurate physics parameters. In general, baselines produce plausible static geometry but exhibit unrealistic dynamics from incorrect material parameters. Our method produces physically consistent deformation and wrinkle patterns throughout the animation.

Figure 4 shows three examples in-the-wild. Even when evaluated on unconstrained real-world imagery, our method produces garments whose simulated dynamics align more closely with the motion observed in the video.

Quantitative Summary. Across both SMS and 4D-Dress, Image2Garment consistently outperforms all baselines in 3D shape metrics. The ablation confirms that this improvement stems specifically from accurate physics parameter estimation rather than garment geometry: fixing ground-truth geometry and varying only the physics parameters still yields 5× lower CD and 1.6× higher IoU compared to randomly sampled materials. This demonstrates that our two-stage factorization — predicting interpretable fabric attributes and mapping them to simulator parameters — is the key driver of improved simulation fidelity.

6 Discussion, Limitations, Future Work, and Conclusion

Discussion. Image2Garment addresses the longstanding challenge of estimating physically accurate, simulation-ready garments from a single in-the-wild image. Our key insight—decomposing the ill-posed image-to-physics mapping through a semantically grounded latent variable of fabric attributes—proves highly effective. Rather than requiring paired image-physics supervision (which does not exist at scale), we leverage the abundant availability of image-to-material data and a modest independently collected material-to-physics dataset.

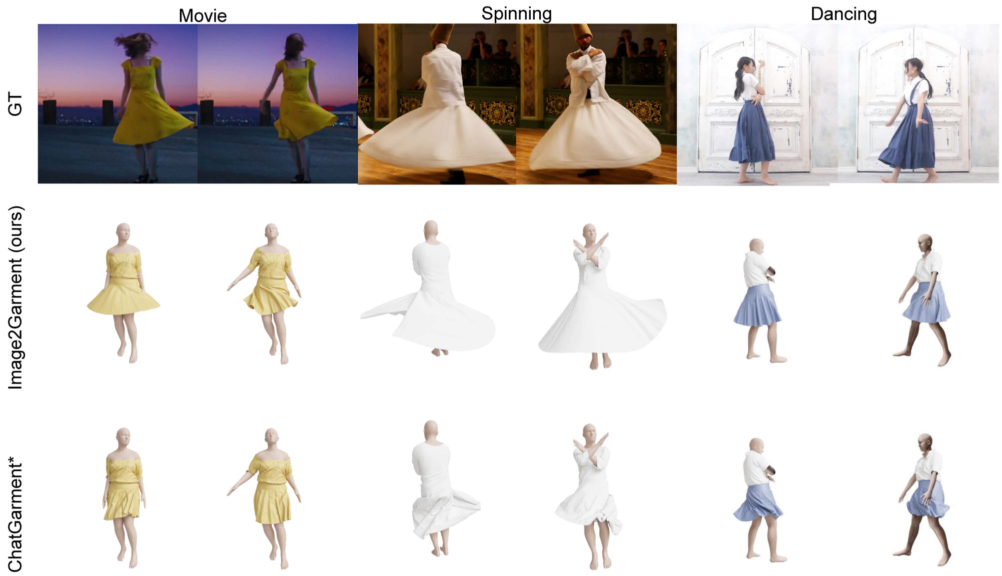

Fig. 4: Qualitative comparison on in-the-wild video. The top row shows the original sequences and the two following rows show the results obtained by our method and ChatGarment*.

This two-stage factorization regularizes the learning problem and dramatically reduces data requirements compared to end-to-end approaches.

Quantitative results on both the synthetic SMS dataset and the real-world 4D-Dress benchmark confirm that accurate physics parameter prediction significantly improves dynamic draping quality. Compared to baselines that use randomly sampled material parameters, our predicted parameters consistently yield simulated garments with lower Chamfer Distance and higher IoU relative to ground-truth cloth geometry. Qualitative results on in-the-wild videos further demonstrate that our method produces temporally consistent, physically plausible garment dynamics across diverse garment types and body motions.

Limitations. Although our model performs well across a diverse range of garments, it has several limitations. First, the method is currently restricted to single-layer clothing; multi-layered or complex garment configurations (e.g., a jacket worn over a shirt) cannot be handled by the present pipeline. Second, garments with prominent accessories or structural elements such as heavy buttons, zippers, or appliqués are not explicitly modeled, which can cause local discrepancies in simulated wrinkle formation near those regions. Finally, our fabric attribute estimation benefits from high-resolution, full-garment context; when only cropped or partially visible garments are available, prediction quality degrades noticeably.

Future Work. The FTAG dataset contains multi-layered garment samples that were filtered out for this work; future research could leverage these samples for layered garment analysis. Finally, integrating accessory-aware physics modeling—e.g., representing localized stiff elements within the cloth simulation—could further improve fidelity for structured garments.

Conclusion. We introduce a feed-forward pipeline that recovers simulation-ready garments from a single image by first predicting structured, interpretable fabric attributes and then mapping them to physically valid simulator parameters. This factorized formulation transforms an otherwise ill-posed inverse problem into two learnable, data-efficient components. Together with our FTAG and T2P datasets, the approach enables accurate estimation of material composition, fabric family, structure type, and full physics parameters directly from in-the-wild images. Extensive quantitative and qualitative results demonstrate state-of-the-art performance in fabric attribute estimation, physics prediction, and dynamic draping quality, while being significantly faster and more scalable than optimization-based approaches. Our results show that semantically grounded material attributes provide an effective foundation for accessible, high-fidelity image-to-simulation pipelines.

References

1. Adobe Creative Cloud: Mixamo (2021), mixamo.com
2. Alldieck, T., Magnor, M., Bhatnagar, B.L., Theobalt, C., Pons-Moll, G.: Learning to reconstruct people in clothing from a single RGB camera. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (jun 2019)
3. Alldieck, T., Pons-Moll, G., Theobalt, C., Magnor, M.: Tex2shape: Detailed full human body geometry from a single image. In: IEEE International Conference on Computer Vision (ICCV). IEEE (2019)
4. Badalyan, A., Selvaraju, P., Becherini, G., Taheri, O., Abrevaya, V.F., Black, M.: Ngl-prompter: Training-free sewing pattern estimation from a single image. arXiv preprint arXiv:2602.20700 (2026)
5. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., Zhong, H., Zhu, Y., Yang, M., Li, Z., Wan, J., Wang, P., Ding, W., Fu, Z., Xu, Y., Ye, J., Zhang, X., Xie, T., Cheng, Z., Zhang, H., Yang, Z., Xu, H., Lin, J.: Qwen2.5-vl technical report (2025), https://arxiv.org/abs/2502.13923
6. Battaglia, P.W., Pascanu, R., Lai, M., Rezende, D., Kavukcuoglu, K.: Interaction networks for learning about objects, relations and physics. In: Advances in Neural Information Processing Systems (NeurIPS) (2016)
7. Bhat, K.S., Twigg, C.D., Hodgins, J.K., Khosla, P.K., Popović, Z., Seitz, S.M.: Estimating cloth simulation parameters from video. In: ACM SIGGRAPH/Eurographics Symposium on Computer Animation (2003)
8. Bhatnagar, B.L., Tiwari, G., Theobalt, C., Pons-Moll, G.: Multi-garment net: Learning to dress 3d people from images. In: IEEE International Conference on Computer Vision (ICCV). IEEE (oct 2019)
9. 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. In: IEEE Conference on Computer Vision and Pattern Recognition (2025)
10. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324, https://doi.org/10.1023/A:1010933404324
11. CLO Virtual Fashion: Marvelous designer (2024), https://www.marvelousdesigner.com/, 3D garment design & cloth simulation software
12. Corona, E., Pumarola, A., Alenyà, G., Pons-Moll, G., Moreno-Noguer, F.: Smplicit: Topology-aware generative model for clothed people. In: CVPR (2021)
13. Feng, X., Huang, W., Xu, W., Wang, H.: Learning-based bending stiffness parameter estimation by a drape tester. ACM Transactions on Graphics (TOG) (2022)
14. Feng, Y., Yang, J., Pollefeys, M., Black, M.J., Bolkart, T.: Capturing and animation of body and clothing from monocular video. In: SIGGRAPH Asia 2022 Conference Papers. SA '22 (2022)
15. Guo, M., Chiang, M.J.Y., Santesteban, I., Sarafianos, N., Chen, H.y., Halimi, O., Božič, A., Saito, S., Wu, J., Liu, C.K., et al.: Pgc: Physics-based gaussian cloth from a single pose. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 21215–21225 (2025)
16. He, K., Yao, K., Zhang, Q., Yu, J., Liu, L., Xu, L.: Dresscode: Autoregressively sewing and generating garments from text guidance (2024), https://arxiv.org/abs/2401.16465
17. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. ICLR (2022)
18. Jiang, B., Hong, Y., Bao, H., Zhang, J.: Selfrecon: Self reconstruction your digital avatar from monocular video. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
19. 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)
20. Korosteleva, M., Lee, S.H.: Neuraltailor: Reconstructing sewing pattern structures from 3d point clouds of garments. ACM Trans. Graph. **41**(4) (2022). https://doi.org/10.1145/3528223.3530179
21. Korosteleva, M., Sorkine-Hornung, O.: GarmentCode: Programming parametric sewing patterns. ACM Transaction on Graphics **42**(6) (2023). https://doi.org/10.1145/3618351, sIGGRAPH ASIA 2023 issue
22. Le, L., Lucas, R., Wang, C., Chen, C., Jayaraman, D., Eaton, E., Liu, L.: Pixie: Fast and generalizable supervised learning of 3d physics from pixels. arXiv preprint arXiv:2508.17437 (2025)
23. 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 (2026), https://arxiv.org/abs/2601.16672
24. Li, R., Dumery, C., Guillard, B., Fua, P.: Garment Recovery with Shape and Deformation Priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024)
25. Li, X., Yu, C., Du, W., Jiang, Y., Xie, T., Chen, Y., Yang, Y., Jiang, C.: Dress-1-to-3: Single image to simulation-ready 3d outfit with diffusion prior and differentiable physics. ACM Transactions on Graphics (TOG) **44**(4), 1–16 (2025)
26. 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 (2024)

27. 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)
28. Moon, G., Nam, H., Shiratori, T., Lee, K.M.: 3d clothed human reconstruction in the wild. In: European Conference on Computer Vision (ECCV) (2022)
29. Nakayama, K., Ackermann, J., Kesdogan, T.L., Zheng, Y., Korosteleva, M., Sorkine-Hornung, O., Guibas, L.J., Yang, G., Wetzstein, G.: Aipparel: A multimodal foundation model for digital garments. In: Proceedings of the Computer Vision and Pattern Recognition Conference (2025)
30. 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: International Conference on 3D Vision 2025 (2025)
31. Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H.: Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In: Proceedings of the IEEE/CVF international conference on computer vision (2019)
32. Tao, Z., Jiang, Y., Chen, Y., Xie, T., Wang, H., Wu, Y., Yang, Y., Kumar, A.S., Tashiro, K., Jiang, C.: Dresswild: Feed-forward pose-agnostic garment sewing pattern generation from in-the-wild images (2026), `https://arxiv.org/abs/2602.16502`
33. 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 Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
34. Xiu, Y., Yang, J., Cao, X., Tzionas, D., Black, M.J.: ECON: Explicit Clothed humans Optimized via Normal integration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2023)
35. Xiu, Y., Yang, J., Tzionas, D., Black, M.J.: ICON: Implicit Clothed humans Obtained from Normals. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13296–13306 (June 2022)
36. Zhang, T., Yu, H.X., Wu, R., Feng, B.Y., Zheng, C., Snavely, N., Wu, J., Freeman, W.T.: Physdreamer: Physics-based interaction with 3d objects via video generation. In: European Conference on Computer Vision. Springer (2024)
37. Zheng, Y., Zhao, Q., Yang, G., Wang, Y., Xiang, D., Dubost, F., Lagun, D., Beeler, T., Tombari, F., Guibas, L., Wetzstein, G.: Physavatar: Learning the physics of dressed 3d avatars from visual observations. In: European Conference on Computer Vision (ECCV) (2024)
38. 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)