# Supplementary Material for FabricTryOn

JUN MA*, Zhejiang Sci-Tech University and Style3D Research, China
QIAN HE†, State Key Lab of CAD&CG, Zhejiang University and Style3D Research, China
GAOFENG HE and HUANG CHEN, Style3D Research, China
CHEN LIU, State Key Lab of CAD&CG, Zhejiang University and Style3D Research, China
XIAOGANG JIN, State Key Lab of CAD&CG, Zhejiang University, China
YIN YANG, University of Utah and Style3D Research, United States of America
HUAMIN WANG, Style3D Research, China

CCS Concepts: • **Computing methodologies** → **Texturing**; **Image processing**.

**ACM Reference Format:**
Jun Ma, Qian He, Gaofeng He, Huang Chen, Chen Liu, Xiaogang Jin, Yin Yang, and Huamin Wang. 2026. Supplementary Material for FabricTryOn. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '26), July 19–23, 2026, Los Angeles, CA, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3799902.3811105

## 1 Implementation Details

In this section, we describe our implementation details, including data curation (Sec. 1.1), model training (Sec. 1.2) and prompts (Sec. 1.3).

### 1.1 Data Curation

The real-world training images are sourced from online e-commerce photoshoots, while the evaluation images are generated with Nano-Banana-Pro. We follow material transfer protocols to evaluate garment re-texturing. Our curated dataset covers diverse garment categories, image layouts and fabric appearances. We avoid paired data from 3D rendering or automated curation, as they may introduce domain gaps or texture misalignment as suboptimal oracles.

We illustrate the fabric scaling and alignment mechanism in Fig. 1. Given a fabric swatch, we first resize it to match the height of the garment image and then apply a scale factor $s$. After scaling and tiling, the fabric region cropped by the garment mask is compared with the garment area using the DISTS metric. We uniformly sample 50 scale factors in $[0, 1]$ and select the one yielding the lowest DISTS score.




Fig. 1. Fabric scaling and alignment using the DISTS metric.

Table 1. Total and valid vote counts in the user studies.

| Criteria | Total Votes | Valid Votes |
|---|---|---|
| Total | 2700 | 2293 |
| Try-on Success | 900 | 756 |
| Fabric Fidelity | 900 | 802 |
| Content Preservation | 900 | 735 |

### 1.2 Model Training

For *fabric removal* and *fabric application*, we adopt Qwen-Image-Edit-2509 as the base model. A LoRA module with rank 16 is trained for 5,000 steps using the AdamW optimizer, with a learning rate of $1 \times 10^{-4}$, weight decay of 0.01, and a batch size of 1. This setting enables the model to robustly remove existing fabric appearance and apply new fabric materials while preserving garment structure.

For *tile augmentation*, we employ FLUX-Fill as the backbone model. Similarly, we train a rank-16 LoRA for 3,000 steps, using AdamW with a learning rate of $1 \times 10^{-4}$, weight decay of 0.01, and a

*This work was conducted during an internship at Style3D Research.
†Project lead and corresponding author.

Authors' Contact Information: Jun Ma, majun88818@163.com, Zhejiang Sci-Tech University and Style3D Research, Hangzhou, China; Qian He, heqianhailie@gmail.com, State Key Lab of CAD&CG, Zhejiang University and Style3D Research, Hangzhou, China; Gaofeng He, hegaofeng@linctex.com; Huang Chen, chenhuang@linctex.com, Style3D Research, Hangzhou, China; Chen Liu, eric.liu@linctex.com, State Key Lab of CAD&CG, Zhejiang University and Style3D Research, Hangzhou, China; Xiaogang Jin, jin@cad.zju.edu.cn, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China; Yin Yang, yangzzzy@gmail.com, University of Utah and Style3D Research, Salt Lake City, Utah, United States of America; Huamin Wang, wanghmin@gmail.com, Style3D Research, Hangzhou, China.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
*SIGGRAPH Conference Papers '26, Los Angeles, CA, USA*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2554-8/2026/07
https://doi.org/10.1145/3799902.3811105

Table 2. User study results with 95% Wilson CIs and $p$-values (two-sided binomial test, $H_0 : p = 0.5$). We collect 100 paired samples for each baseline method, yielding 900 evaluations across nine methods. Values indicate the percentage of votes in which each baseline is preferred over ours. "0.0" for SD methods indicates zero preference votes. "–" for Ours denotes self-comparison is not applicable.

| Method | Method Configuration | | | Try-on Success | | Fabric Fidelity | | Content Preservation | |
|---|---|---|---|---|---|---|---|---|---|
| | Base Model | Strategy | Training | 95% CI | $p$-value | 95% CI | $p$-value | 95% CI | $p$-value |
| MatSwap | SD v1.5 | - | Full model | 0.0–3.7 | $< 0.001$ | 0.0–3.7 | $< 0.001$ | 0.0–3.7 | $< 0.001$ |
| ZeST | SDXL | - | - | 0.0–3.7 | $< 0.001$ | 0.0–3.7 | $< 0.001$ | 0.0–3.7 | $< 0.001$ |
| MaterialFusion | SD v1.5 | - | - | 0.0–3.7 | $< 0.001$ | 0.0–3.7 | $< 0.001$ | 0.0–3.7 | $< 0.001$ |
| | GPT-Image-1.5 | 2-Step | - | 6.0–20.0 | $< 0.001$ | 9.0–23.2 | $< 0.001$ | 3.0–13.2 | $< 0.001$ |
| | Nano-Banana-Pro | 2-Step | - | 15.6–33.8 | $< 0.001$ | 29.2–49.5 | 0.0503 | 14.7–32.8 | $< 0.001$ |
| | Kontext | 2-Step | LoRA | 8.9–25.7 | $< 0.001$ | 7.6–21.4 | $< 0.001$ | 7.2–22.3 | $< 0.001$ |
| | Qwen-Image-Edit | 1-Step | - | 2.5–12.8 | $< 0.001$ | 5.5–18.5 | $< 0.001$ | 3.2–16.1 | $< 0.001$ |
| | Qwen-Image-Edit | 2-Step | - | 3.2–14.2 | $< 0.001$ | 2.5–12.9 | $< 0.001$ | 3.8–16.8 | $< 0.001$ |
| | Qwen-Image-Edit | 1-Step | LoRA | 23.6–49.0 | 0.0489 | 20.6–43.8 | 0.0054 | 27.0–54.5 | 0.2327 |
| **Ours** | **Qwen-Image-Edit** | **2-Step** | **LoRA** | – | – | – | – | – | – |

batch size of 1. This module focuses on enriching local fabric details and improving material continuity under partial editing scenarios.

### 1.3 Prompts

*VLM-based Classification.* You are given an image of garments, either flat-laid or worn by a model. Identify the single most complete garment unit shown. A garment unit may be one standalone garment or a coordinated set designed to be worn together and made from the same fabric (for example, a pajama set). Only treat a set as one unit if all pieces are fully visible, structurally complete, and clearly share the same fabric texture and material. If multiple garment units appear in the image, choose the most complete one. Respond with only one standardized category and nothing else.

*VLM-based filtering for fabric extraction.* Determine whether the fabric patch can be reliably used as a representative sample of the *garment_category* shown. The patch should clearly preserve the garment's fabric texture, material structure, and pattern details. If the patch is overly blurred, low quality, fails to reflect the original fabric texture, or contains mismatched areas or visual artifacts, respond "No". Otherwise, respond "Yes". Respond with only "Yes" or "No".

*VLM-based filtering for fabric removal.* Determine whether the fabric removal result satisfies the requirements for the *garment_category* shown. All fabric textures, materials, patterns, and stripes should be fully removed, leaving the garment smooth, plain, and uniformly mid-gray. The original garment style, shape, structural details, and lighting should be preserved and consistent with the input image. If any fabric texture, pattern, stripe, or material detail remains, or if the garment's style, geometry, or lighting differs from the original image, respond "No". Otherwise, respond "Yes". Respond with only "Yes" or "No".

*Fabric Extraction.* Generate a seamless, tileable fabric texture that matches the fabric used in the *garment_category*. The fabric should be laid perfectly flat and fill the entire image edge-to-edge, with no background, borders, or empty space. The texture must repeat continuously in all directions with no visible seams or breaks. Accurately match the garment's color, fabric texture, material properties, weave or knit structure, and thickness.

*Fabric Removal.* Remove all materials, prints, and stripes from the *garment_category*, converting it into a smooth, uniform gray surface. The garment should resemble a 3D-rendered white model with no texture or material details, while preserving its original design, shape, and structure. Keep the background unchanged and leave all other clothing untouched.

*Fabric Application.* Replace the material of the *garment_category* with the provided fabric, making it appear as though the garment was originally made from this material. The fabric should drape and deform naturally according to the garment's structure, showing realistic folds, shading, and texture variation, rather than looking flat or pasted on. Preserve the original garment's style and shape, and maintain the fabric's texture scale and pattern density. Keep the background unchanged and leave all other clothing untouched.

## 2 User Study

We present the implementation details of the user study and the statistical significance of the results. Fig. 19 illustrates our user study interface. For each question, users select a preferred method or indicate *no clear preference*, which is counted as an invalid vote. The total and valid vote counts are reported in Tab. 1. As shown in Tab. 2, we present the results with 95% Wilson Confidence Intervals (CIs) for binomial proportions, and assess statistical significance using a two-sided binomial test with the null hypothesis of equal preference ($H_0 : p = 0.5$).

The user study shows a clear preference for our method over most baselines across all three criteria. For methods such as MatSwap, ZeST, and MaterialFusion, the 95% Wilson confidence intervals remain below 4% with $p$-values $< 0.001$, indicating statistically significant and near-zero preference compared to ours. Similarly, Qwen variants and Kontext-2Step-LoRA exhibit low preference rates with strong statistical significance, confirming consistent user favor toward our approach. More competitive methods, including GPT-Image-1.5 and Nano-Banana-Pro, achieve higher preference ranges;




Fig. 2. Results of our method across different training dataset sizes. Improvements in material removal, region control, and fabric fidelity quickly plateau.

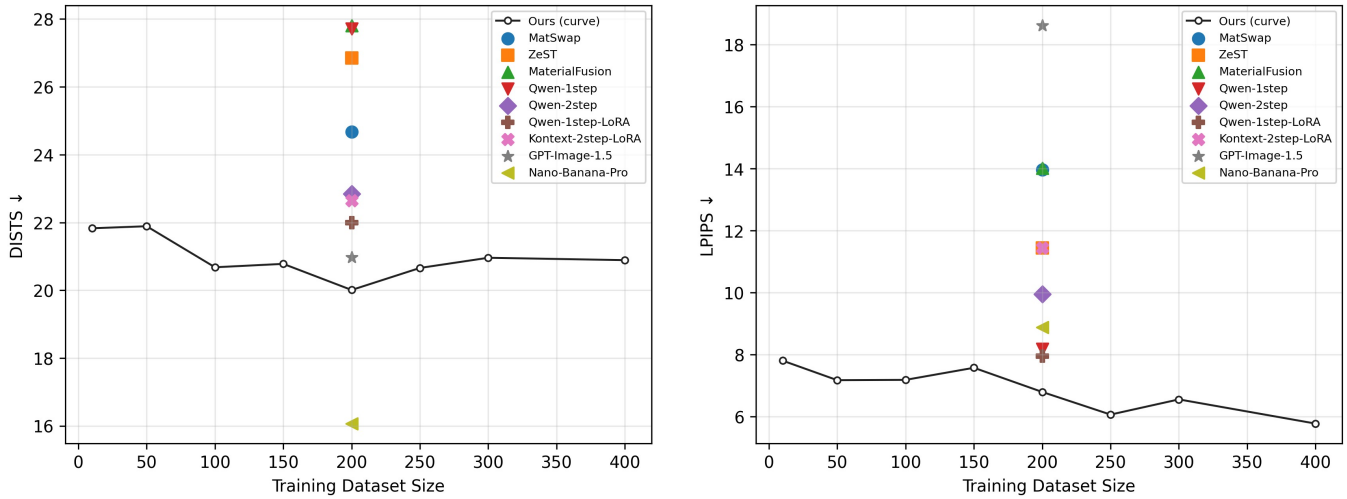


Fig. 3. Ablation study on training dataset size (metrics scaled by ×100).

however, their results remain significantly below chance ($p < 0.001$) in most cases, suggesting that our method still outperforms them reliably. Nano-Banana-Pro shows borderline significance in fabric fidelity ($p = 0.0503$), indicating comparable performance on this specific aspect due to the fabric overlaying phenomenon. Qwen-1Step-LoRA is the closest competitor, with confidence intervals approaching 50%. While it remains worse than our method in try-on success ($p = 0.0489$) and fabric fidelity ($p = 0.0054$), the difference in content preservation is not statistically significant ($p = 0.2327$), suggesting comparable performance in maintaining garment structure. Overall, the results demonstrate that our method achieves statistically significant advantages over most baselines, with only marginal gaps observed in specific criteria for the strongest competing methods.

## 3 Ablation on Dataset Size

In this work, we aim to activate garment re-texturing capability in MMDiTs using a small curated dataset, and conducted ablation studies on the number of training samples (10-400). All models are trained for 5000 steps. As shown in Fig. 3 and Fig. 2, quantitative performance saturates quickly, with improvements in material removal, region control, and fabric fidelity rapidly plateauing. Ours (Train-200) generalizes well to in-the-wild scenarios (Fig. 15), demonstrating the strong learning efficiency of the proposed framework.

## 4 Analysis of Method Components

In this section, we analyze key components of our method, including VLM-based filtering in the data curation pipeline and comparisons between our material-normalized representation and alternatives such as Nano-Banana-Pro outputs and the irradiance representation in RGB↔X [Zeng et al. 2024].

### 4.1 VLM-based Filtering

We adopt Qwen2.5-VL to filter out obvious failure generations in our data curation pipeline. For small training sets, manual verification is efficient to ensure data quality. Additionally, LoRA training smooths minor material variations, making the framework tolerant to slightly inaccurate masks or material-normalized images. If scaling to larger datasets, stronger VLMs such as Qwen-3.5 would further improve accuracy.

We evaluate the filtering results on the Nano-Banana-Pro fabric removal outputs using the evaluation dataset, where Qwen2.5-VL achieves a passing ratio of 80%, compared to 63.3% for manual human filtering. The higher passing ratio indicates that VLM-based filtering is more permissive and may allow some failure cases to pass through.

Typical failure cases of VLM-based filtering are further analyzed in Fig. 4. For fabric removal results, the VLM tends to fail on samples with residual fine-grained textures or incorrect material changes, while it can reliably detect structural or semantic deviations such as style changes and unintended modifications on non-target regions. For fabric extraction results, the VLM is effective at identifying semantic errors, but may fail to detect subtle low-level defects, such as internal seam artifacts and slight color shifts.

### 4.2 Material Normalization

In Fig. 5, we compare our material-normalized results with those produced by the Qwen-Image-Edit base model and Nano-Banana-Pro. Without fine-tuning, Qwen-Image-Edit fails to fully remove original fabrics (e.g., blue denim in Case 1 and knitted structures in Case 3) and may simultaneously modify non-target region or alter garment shape (Case 2). Nano-Banana-Pro exhibits similar variations, motivating the filtering stage in our data curation pipeline. After fine-tuning on our curated dataset, our method produces more stable and consistent material-normalized garment images, and these

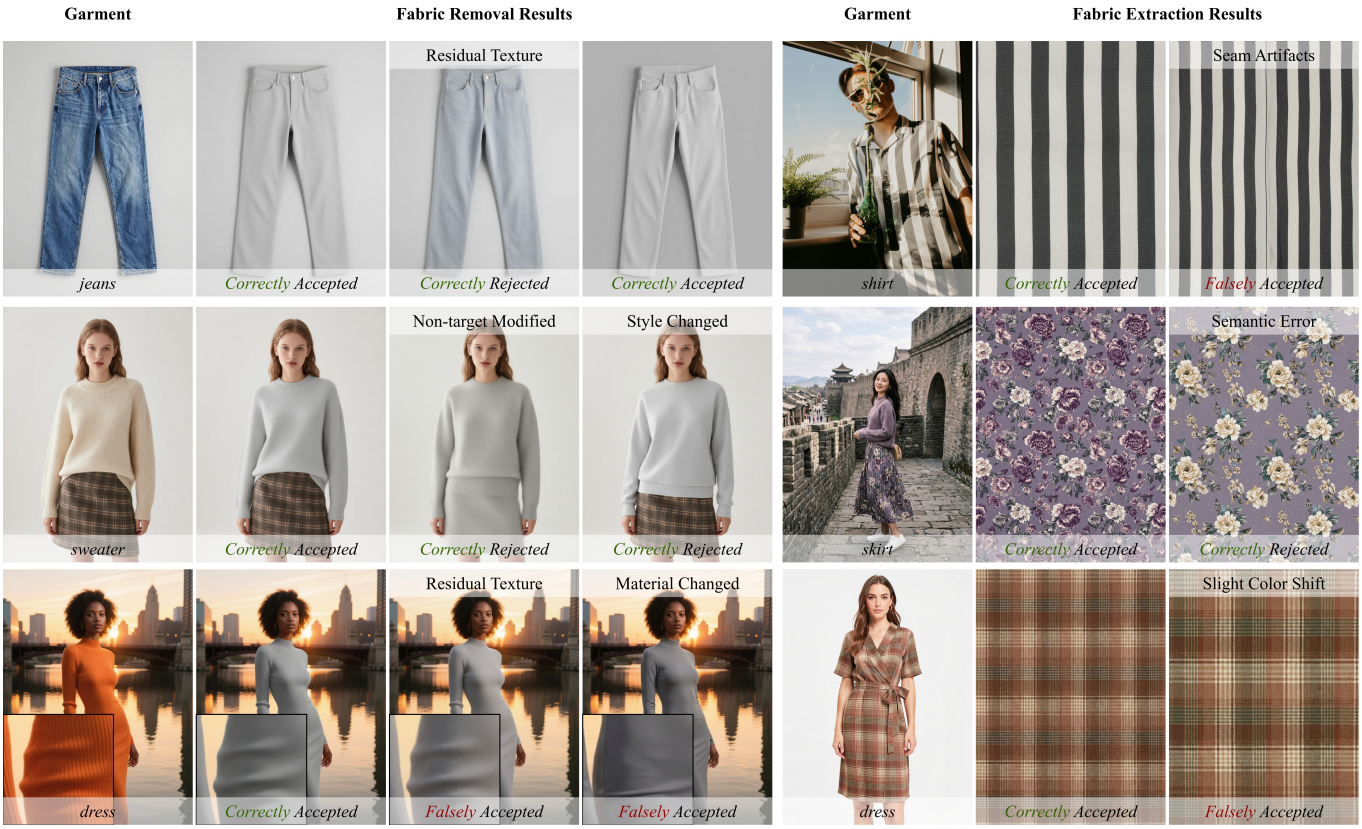


Fig. 4. Visual examples of VLM-based filtering. We present diverse generated examples to the filtering module, illustrating both its successes and failure cases. Image credits (via Unsplash): Row 1, Column 5 © Nathan McDine.

Table 3. Quantitative comparison of our material-normalized results with Nano-Banana-Pro outputs and RGB-X Irradiance.

| Method for Fabric Removal | Qwen-VL-pass-ratio | Human-pass-ratio | TryOn-DISTS ↓ | TryOn-LPIPS ↓ | Rec.-PSNR ↑ | Rec.-LPIPS ↓ |
|---|---|---|---|---|---|---|
| Nano-Banana-Pro | 80.0% | 63.3% | 17.83 ± 4.56 | 10.62 ± 6.36 | – | – |
| RGBX-irradiance | – | – | 19.90 ± 4.08 | 8.29 ± 4.88 | 14.03 ± 1.98 | 47.34 ± 13.63 |
| Ours | 93.3% | 93.3% | 20.01 ± 4.28 | 6.80 ± 3.56 | 13.82 ± 1.89 | 46.66 ± 13.57 |

advantages further enhance the quality of subsequent re-texturing results.

*Comparison with Nano-Banana-Pro.* Our fabric removal LoRA achieves a higher Qwen-VL-pass-ratio and Human-pass-ratio than Nano-Banana-Pro, as shown in Tab. 3. Replacing it with Nano-Banana-Pro in our two-step re-texturing pipeline degrades geometric details (higher TryOn-LPIPS) and fabric-garment integration (lower TryOn-DISTS), similar to Nano-Banana-Pro (2-step) in the main paper. This improvement arises because LoRA learns a filtered, high-quality subset of the Nano-Banana-Pro's outputs, effectively producing a sharpened approximation of the best outputs rather than modeling the full stochastic distribution.

*Comparison with RGB↔X Irradiance.* Our material-normalized image is designed for garment re-texturing rather than generic physics-based rendering [Kocsis et al. 2025, 2024], conceptually differing from traditional components like shading (varied materials) [Careaga and Aksoy 2023] or irradiance (inaccurate geometry). We also provide a quantitative comparison with RGBX-irradiance in Tab. 3. For garment re-texturing (Fig. 6), replacing our material-normalized garment with RGBX-irradiance produces worse geometric preservation (higher TryOn-LPIPS) while comparable fabric fidelity (close TryOn-DISTS). However, it shows weaker generalization to in-the-wild images compared to our method and may retain residual textures or introduce overexposure artifacts, as illustrated in Case 3 and Case 4 of Fig. 6. For X-to-RGB reconstruction (Fig. 7), using our material-normalized garment as irradiance achieves comparable results (close Rec.-PSNR and Rec.-LPIPS). Note that the results are strongly influenced by the robustness and limitations of the downstream model.

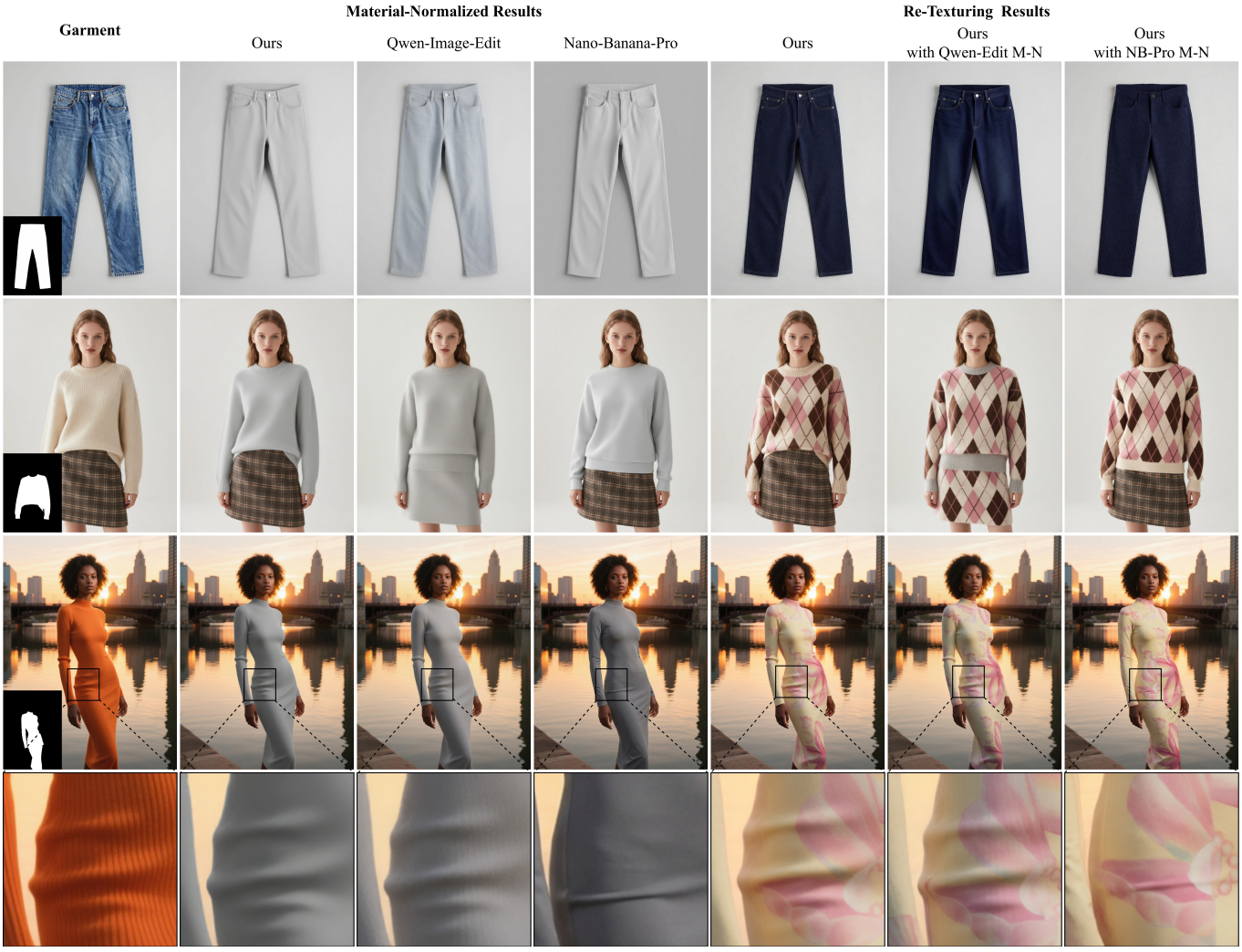


Fig. 5. Comparison on material-normalized results and subsequent re-texturing results.

## 5 Further Comparisons on Garment Re-Texturing

In this section, we present further comparisons on garment re-texturing, including evaluations against one-step methods (Sec. 5.1), Nano-Banana-Pro under different input settings (Sec. 5.2), and 3D garment texturing (Sec. 5.3).

### 5.1 Comparison with One-Step Methods

In Fig. 8, we compare our results with state-of-the-art image editing models using one-step generation, where all methods are given the same original garment image and fabric exemplar. Qwen transfers little to no fabric, often reproducing the original garment. Both GPT and Banana preserve specular reflectance from the original fabric in Case 1. GPT further alters dress folds in Case 2 and introduces color shifts in human faces in Cases 3–5. Banana exhibits severe artifacts, frequently overlaying the fabric without proper integration. In contrast, our Qwen-based method achieves superior disentanglement of fabric material and garment geometry, and more coherent fabric integration, through fine-tuning on a small curated dataset. The initial capability of our base model (Qwen-Image-Edit) compared to stronger baselines highlights the effectiveness of our two-stage learning.

### 5.2 Comparison with Nano-Banana-Pro

Since Nano-Banana-Pro is the strongest image editing model for fabric manipulation, we conduct a comprehensive analysis under different input variants. As shown in Fig. 9, given a garment image and a fully tiled fabric, Nano-Banana-Pro produces re-textured garments with inconsistent pattern density (Case 1, left). When using a masked fabric input (one-step setting as in Fig. 8), unintended material blending between the new and original fabrics remains, and the new fabric often appears overlaid without proper deformation (Case 2, left). Even when provided with a material-normalized

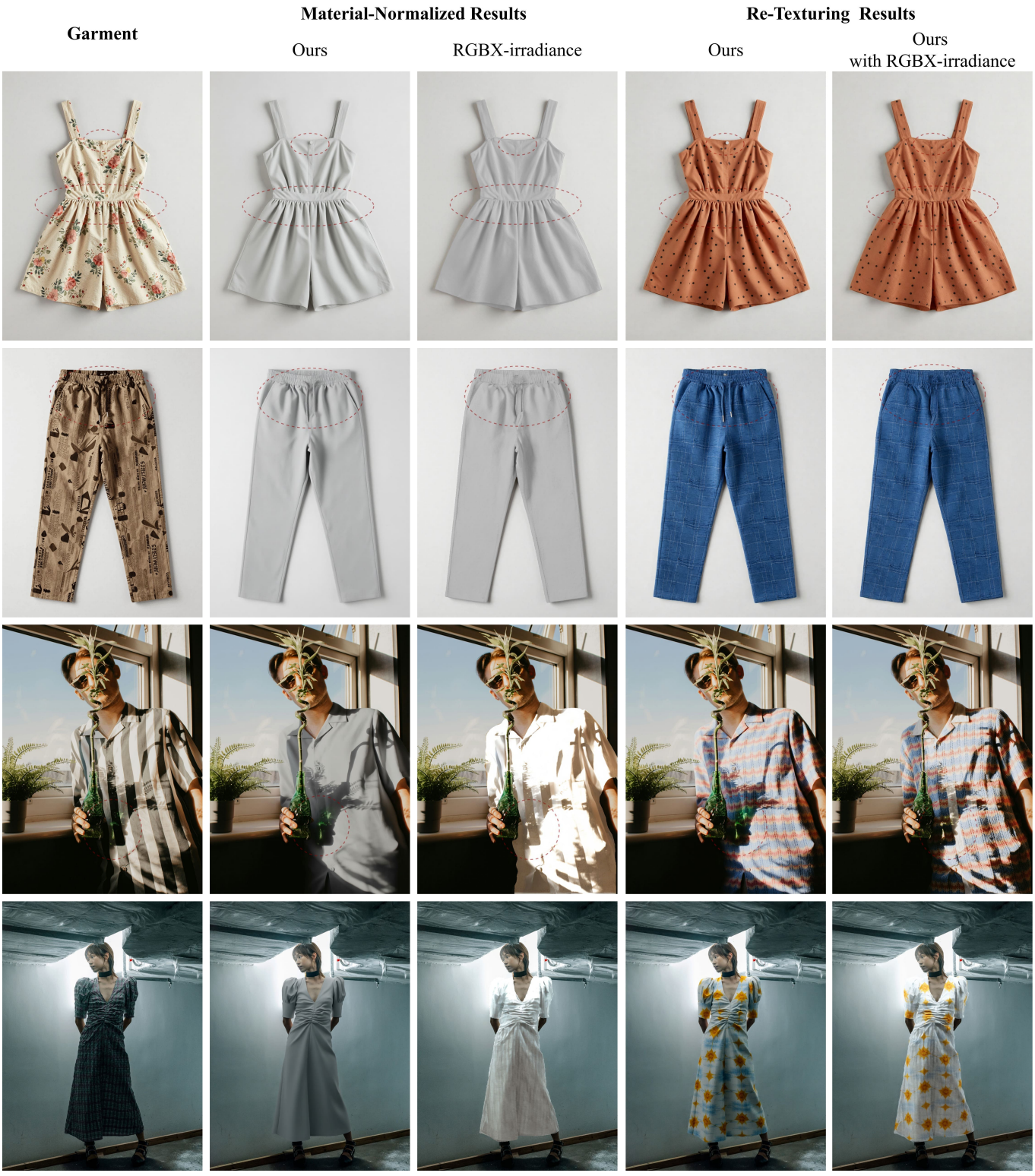


Fig. 6. Comparison of our material-normalized results with RGB↔X irradiance in garment re-texturing. For RGBX-irradiance, we replace the foreground garment region of our material-normalized image with the irradiance predicted by RGB↔X. Image credits (via Unsplash): Row 3, Column 1 © Nathan McDine; and Row 4, Column 1 © Dynamic Wang.

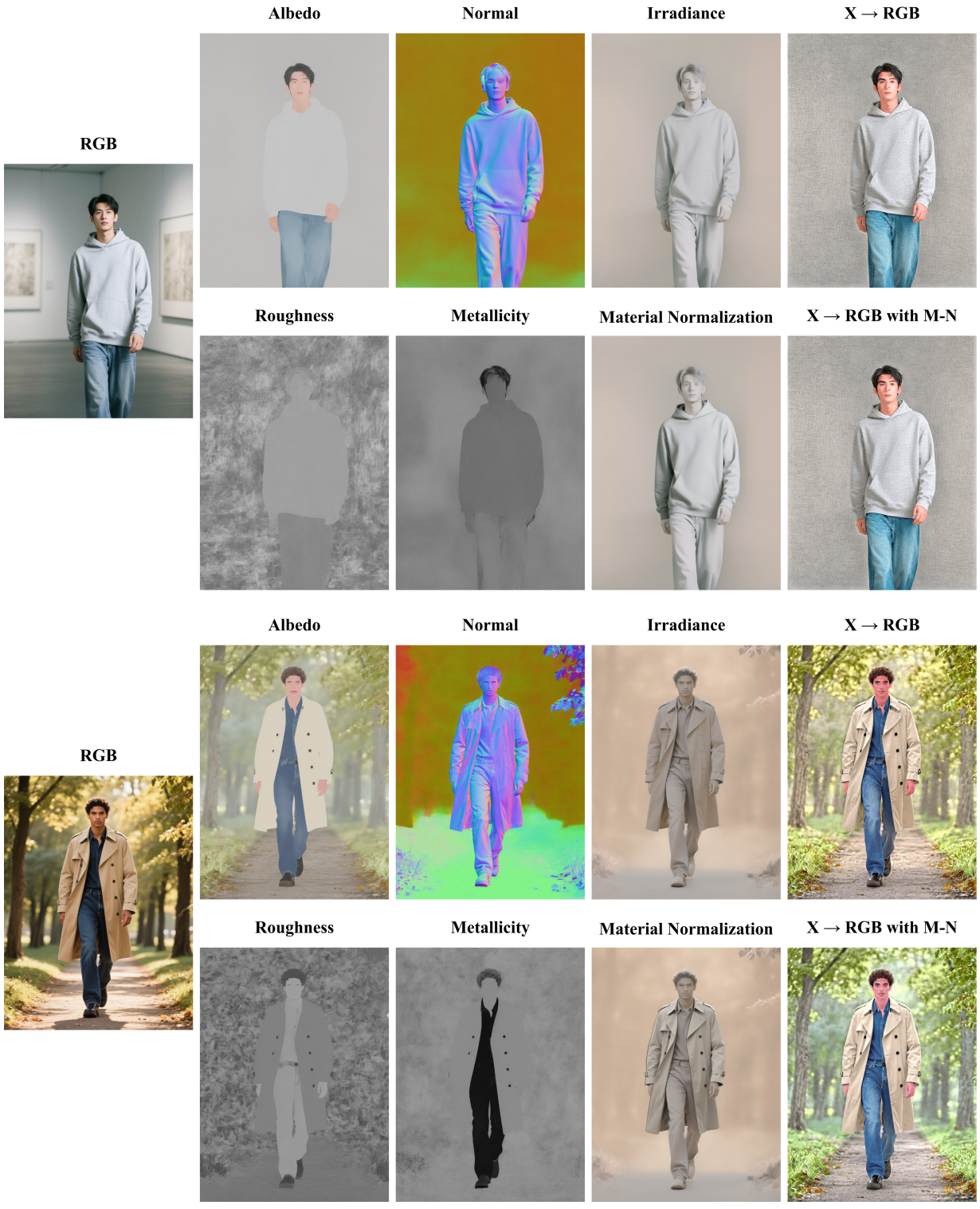


Fig. 7. Comparison of our material-normalized results with RGB↔X irradiance in X-to-RGB reconstruction. For Material Normalization, we replace the foreground garment region of RBGX-irradiance with our material-normalized result.

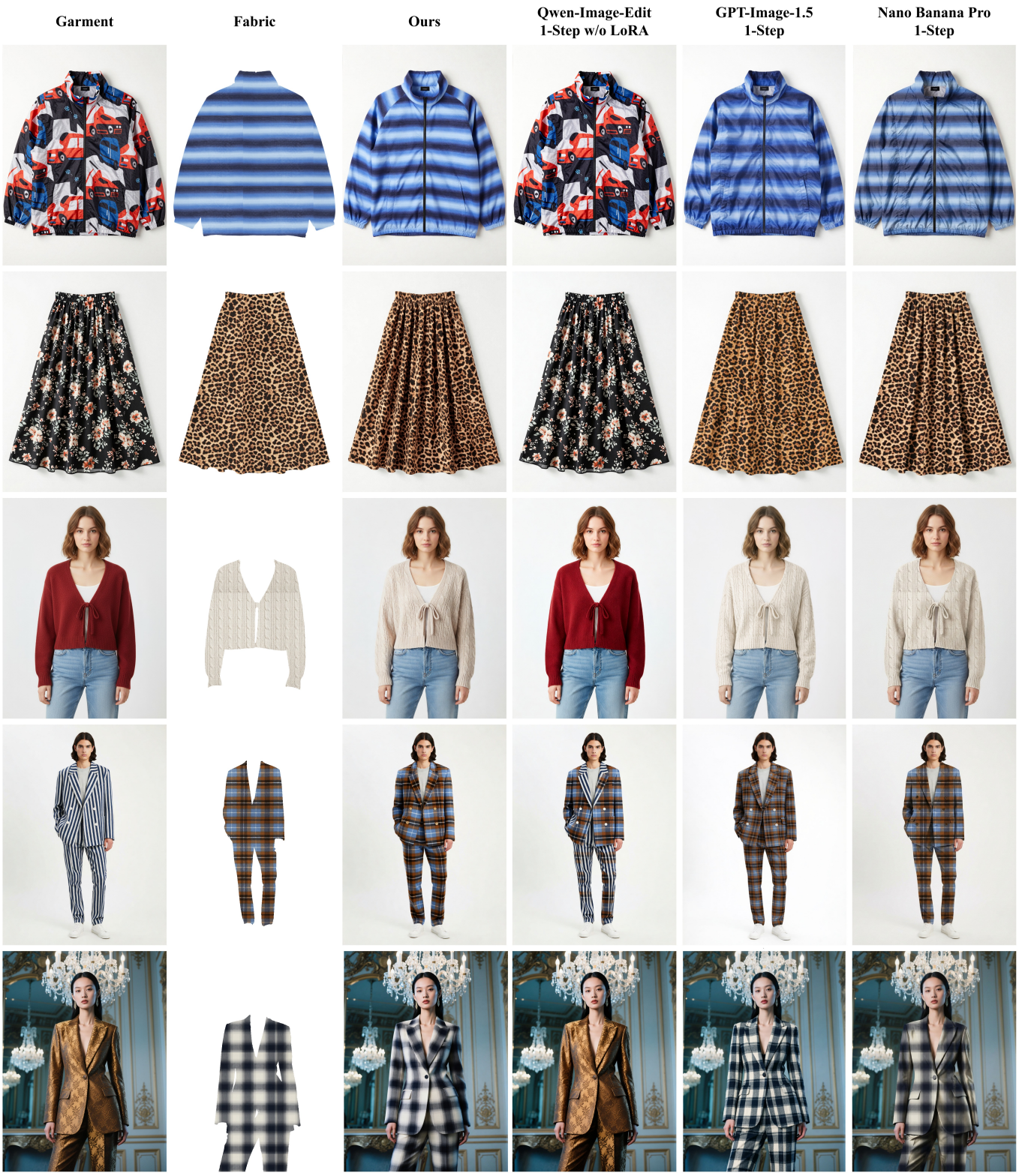


Fig. 8. Additional ablation results with one-step variants.

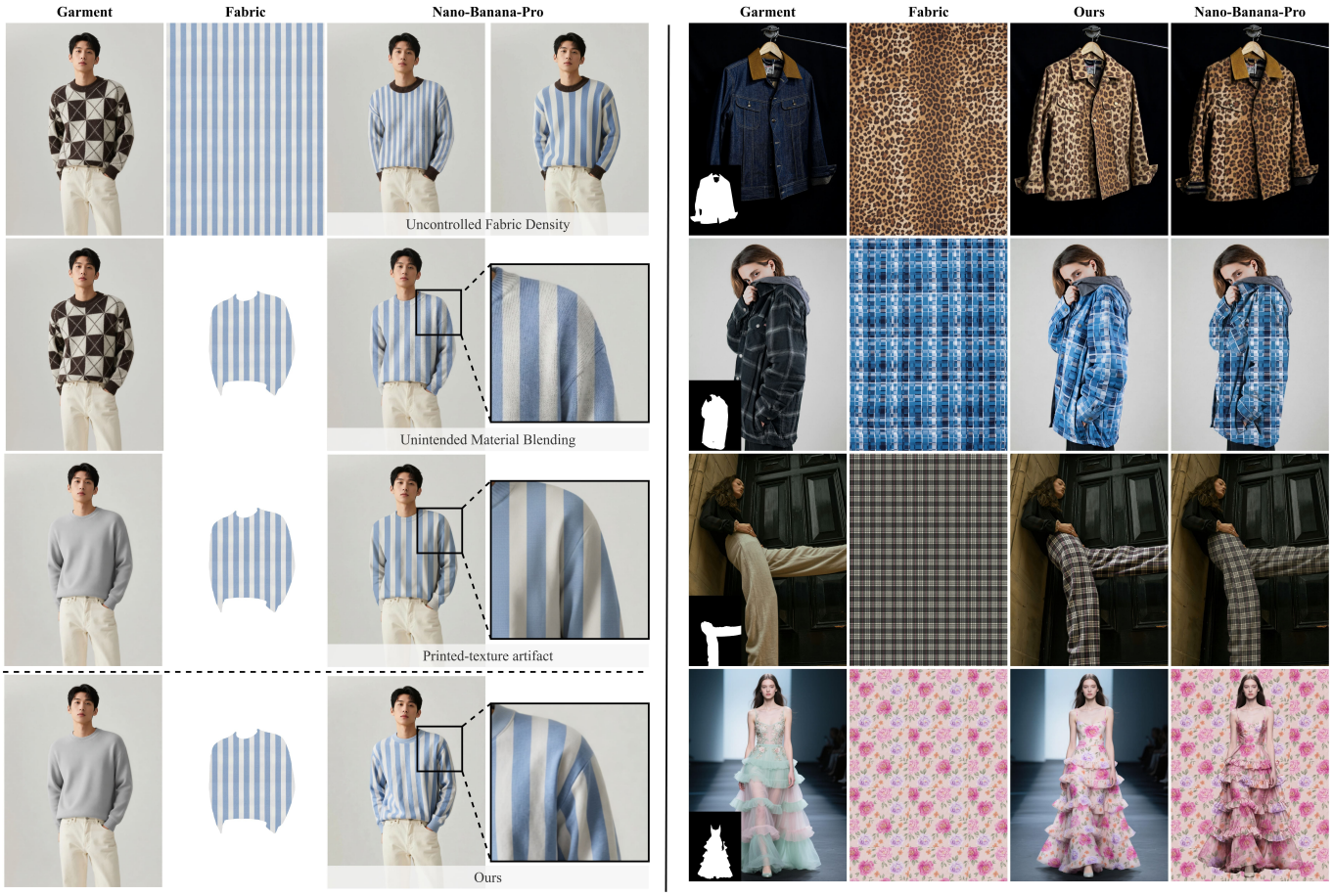


Fig. 9. Comparison with Nano-Banana-Pro on garment re-texturing. Left: characteristics of Nano-Banana-Pro outputs under different input variants. Right: results using default fabric images as inputs, analogous to material transfer methods. **Ours** refers to results of our full two-step re-texturing pipeline. Image credits (via Unsplash): Row 1, Column 5 © Caio Coelho; Row 2, Column 5 © Ali Choubin; and Row 3 Column 5 © Korie Cull.

garment input, printed-texture artifacts persist (Case 3, left). In contrast, our full method produces well-integrated re-textured garments without interference from the original fabric.

Since we have already compared Nano-Banana-Pro in one-step, two-step, and hybrid settings with our fabric application module (Sec. 4.2), we further present four examples using default fabric images as inputs (right of Fig. 9), analogous to material transfer methods. The results exhibit persistent overlay artifacts and poor integration across different garment types, highlighting the need for a specialized approach to garment re-texturing.

## 5.3 Comparison with 3D Garment Texturing

Although our work focuses on image-based garment re-texturing, we also compare against 3D texturing and rendering using 3D garment models to highlight the advantages of generative models. As shown in Fig. 10, we first render white 3D garments and their textured counterparts using the fabric swatches. We then treat the white initial renderings as input garment images and apply our method to generate re-textured results. Compared to conventional 3D UV mapping, our generative approach better captures knitwear textures and produces richer tactile appearance that is difficult to achieve with standard PBR shading. Additionally, localized seams and surface irregularities, as well as the ribbed structures at the cuffs and collar, are automatically synthesized and blended by the model. In contrast, achieving comparable detail with traditional 3D texturing requires substantial manual effort.

## 6 Comparison on Diverse Objects

We present additional qualitative comparisons with material transfer methods on diverse object categories in Fig. 12. Our method generalizes well to objects with varied shapes and deformations. While MatSwap, MaterialFusion, and ZeST produce more plausible results on these objects than on garments due to reduced fine-grained geometric deformations, their outputs remain inferior to ours.




Fig. 10. Comparison with 3D Garment Texturing.

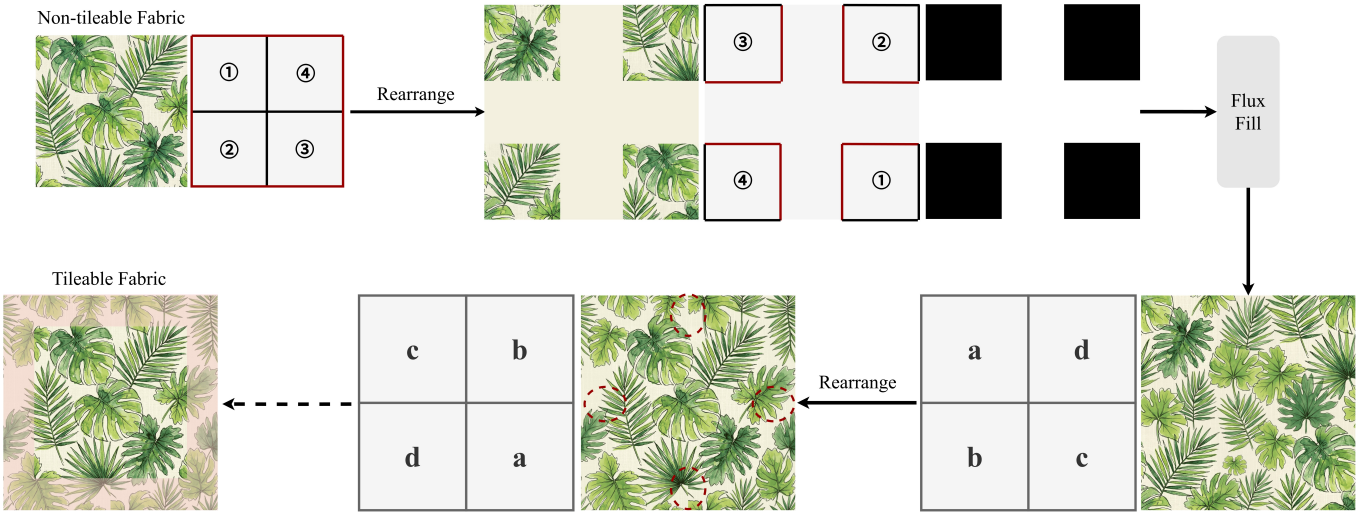


Fig. 11. Adaptation of ControlMat to our tile augmentation pipeline. For a fair comparison, we employ a fixed cross-shaped mask with the same area as our diamond mask. The red dashed ellipses indicate potential seams introduced after generation and rearrangement.

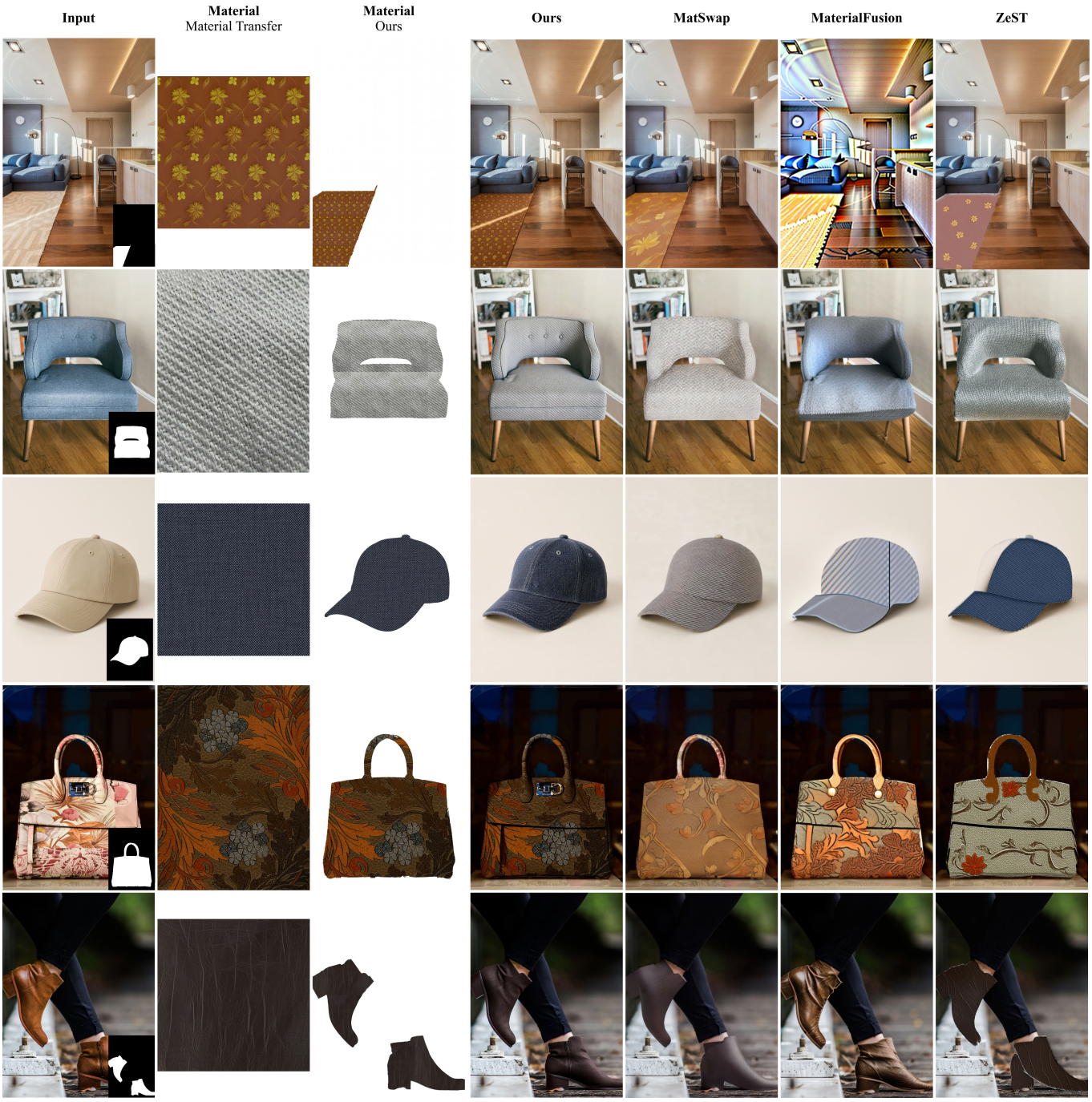


Fig. 12. Comparison with material transfer methods on diverse object categories. Our model uses a masked fabric exemplar, while baselines receive fully tiled fabrics, as white background regions degrade IP-Adapter feature injection in their pipelines. Image credits (via Unsplash unless otherwise noted): Row 1, Column 1 © Unsplash; Row 1, Column 2 © MatSynth [Vecchio and Deschaintre 2024]; Row 2, Column 1 © Kari Shea; Row 2, Column 2 © Second Breakfast; Row 4, Column 1 © Arno Senoner; Row 4, Column 2 © The New York Public Library; and Row 5, Column 1 © Verity Sanders.

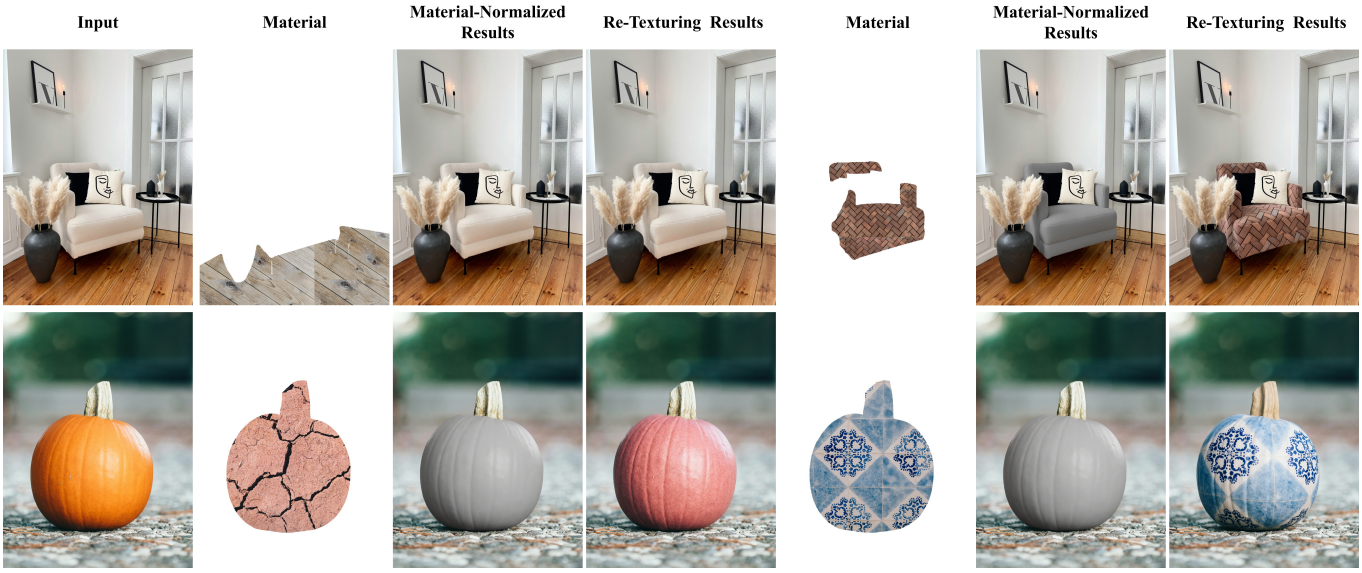


Fig. 13. Application of our trained re-texturing pipeline on non-fabric materials. Image credits (via Unsplash): Row 1, Column 1 © volant; Row 1, Column 2 © Alexandre Boucher; Row 1, Column 5 © Lluvia Morales; Row 2, Column 1 © Griffin Wooldridge; Row 2, Column 2 © Peaky Frames; and Row 2, Column 5 © Annie Spratt.

## 7 Extended Comparison on Tile Augmentation

We conduct an extended comparison on tile augmentation. ControlMat [Vecchio et al. 2024] also adopts an inpainting-based framework, but it is neither directly applicable nor a superior solution. Its noise-rolling strategy is effective for dense, low-level repetitive patterns but disrupts structured semantic features (see their limitation, Fig. 15), as shifting inpainting regions across denoising steps breaks contextual consistency. This design also complicates adaptation from UNets to DiTs due to image tokenization and positional encoding. For comparison, we fix the noise-rolling position as a cross-shaped mask, analogous to our diamond region, and apply inpainting with FLUX-Fill (Fig. 11). For fairness, the cross mask has a width of $\sqrt{2}-1$ times the input fabric edge and matches the area of our diamond mask. However, this cross-based inpainting still produces visible seams at the endpoints, and balancing inpainting width with generalization across diverse textures remains challenging.

We present additional qualitative comparisons in Fig. 14. We focus on fabric fidelity and tileability without enforcing strict resolution alignment, as different methods produce tilable units at varying resolutions. FLUX-Fill with the cross mask produces seams in the generated regions (Case 1) and may introduce severe artifacts due to thin boundary constraints (Case 2). Content-Aware-Tiles with wang/dual tiling significantly alters the original fabric textures. Offset-Blend, Edge-Mirroring, and Image-Quilting all disrupt semantic patterns and introduce noticeable deviations from the input appearance. In contrast, our method achieves a strong balance between tileability and semantic coherence.

We further conduct a quantitative analysis of FLUX-Fill–based inpainting methods under different masking strategies, by comparing the generated (inpainted) regions with the input fabrics, as summarized in Tab. 4. The CLIP-scores indicate semantic consistency in the inpainted results and highlight the benefits of our diamond mask arrangement and LoRA fine-tuning. We also compute a high-frequency (HF) energy ratio to measure fine-grained fabric details. Specifically, for a region $R$, the high-frequency energy is defined as:

$$E(R) = \frac{1}{|R|} \sum_{p \in R} |\Delta I(p)|, \tag{1}$$

where $\Delta I(p)$ denotes the Laplacian response at pixel $p$. The relative HF ratio is then defined as:

$$\text{Ratio}(R) = \frac{E(R_{\text{gen}})}{E(R_{\text{ori}})}. \tag{2}$$

The original region $R_{\text{ori}}$ corresponds to the input fabric image, while $R_{\text{gen}}$ denotes the inpainted region. A higher ratio indicates a closer resemblance in fine-grained structures. Our improvement over FLUX-Fill (diamond) demonstrates the effectiveness of LoRA fine-tuning in enhancing high-frequency fabric details.

Table 4. Quantitative comparison between generated (inpainted) regions and input fabric images (scaled by ×100).

| Method | CLIP-Score ↑ | HF Energy Ratio ↑ |
|---|---|---|
| **Ours (diamond)** | **94.04 ± 1.59** | **84.48 ± 4.97** |
| FLUX-Fill (diamond) | 93.83 ± 1.93 | 79.19 ± 4.73 |
| FLUX-Fill (cross) | 92.67 ± 3.45 | 82.69 ± 6.39 |

## 8 More Results of Our Method

In this section, we demonstrate additional qualitative results of our method, including garment re-texturing (Sec. 8.1) and tile augmentation (Sec. 8.2).

### 8.1 Garment Re-Texturing

We present additional garment re-texturing results of our method in Fig. 15, Fig. 16 and Fig. 17.

### 8.2 Tile Augmentation

We present additional tile augmentation results in Fig. 18. Our tile augmentation module generates seamless fabric tiles that closely match the input fabric, preserving high-frequency textures and woven structures.

## 9 Application to Non-Fabric Materials

We apply our trained re-texturing pipeline to non-garment objects and non-fabric materials, as illustrated in Fig. 13. For objects and materials with a large domain gap from garments and fabrics (e.g., the wooden floor in Case 1), the material-normalization module may fail, yielding results nearly identical to the input. For objects such as the pumpkin in Case 3, integration with mud can be unsatisfactory. In contrast, for materials such as brick or ceramic tile, both the sofa and the pumpkin integrate well with our pipeline, though some fine-grained reflectance properties may be lost.

## 10 Ethics

Our work enables realistic and accessible garment editing, which may introduce risks of misuse, such as deceptive product advertising or manipulation of the visual message conveyed by an image. To mitigate these concerns, we recommend practical safeguards including watermarking edited outputs, attaching provenance metadata to indicate model-based modifications, and enforcing disclosure policies in downstream applications. Platform-level moderation and clear usage guidelines can further help reduce potential misuse.

We also acknowledge potential dataset bias. The training data may not fully capture the diversity of body types, skin tones, and clothing styles, which could affect model performance and fairness across different groups. Future iterations will incorporate broader dataset diversity and systematic bias assessments to ensure more inclusive and reliable results.

## Acknowledgments

Xiaogang Jin was supported by the National Natural Science Foundation of China (Grants: U25A20440, 62472373).

## References

Chris Careaga and Yağız Aksoy. 2023. Intrinsic image decomposition via ordinal shading. *ACM Transactions on Graphics* 43, 1 (2023), 1–24.

Peter Kocsis, Lukas Höllein, and Matthias Nießner. 2025. Intrinsix: High-quality pbr generation using image priors. *arXiv preprint arXiv:2504.01008* (2025).

Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. 2024. Intrinsic image diffusion for indoor single-view material estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5198–5208.

Giuseppe Vecchio and Valentin Deschaintre. 2024. Matsynth: A modern pbr materials dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22109–22118.

Giuseppe Vecchio, Rosalie Martin, Arthur Roullier, Adrien Kaiser, Romain Rouffet, Valentin Deschaintre, and Tamy Boubekeur. 2024. Controlmat: a controlled generative approach to material capture. *ACM Transactions on Graphics* 43, 5 (2024), 1–17.

Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. 2024. RGB↔X: Image decomposition and synthesis using material- and lighting-aware diffusion models. In *ACM SIGGRAPH 2024 Conference Papers*. Article 75, 11 pages.

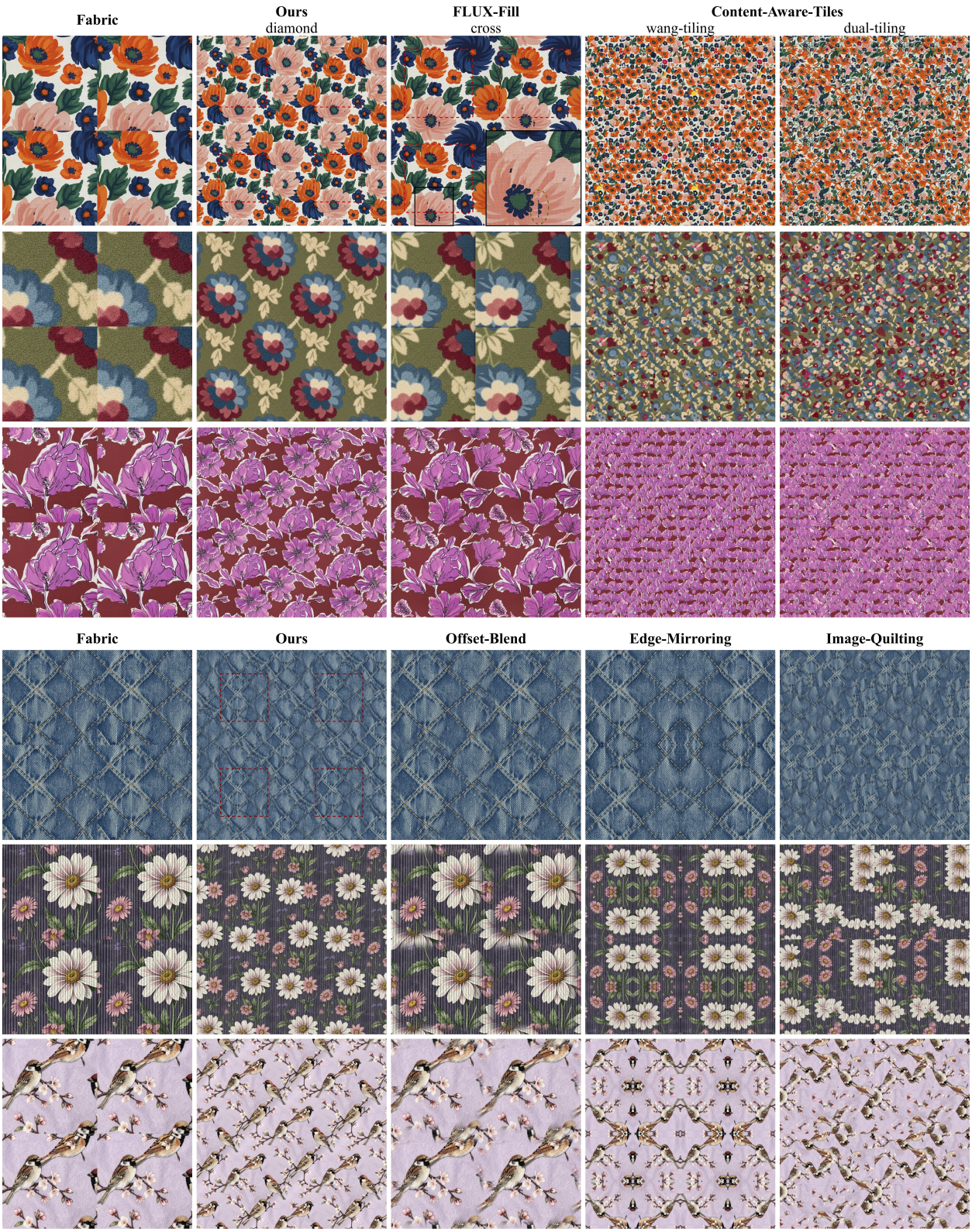


Fig. 14. Comparison on tile augmentation. Column 1 shows the $2 \times 2$ tiled input fabrics, and Columns 2–5 show tiled repetitive units from different methods. Red dashed boxes indicate regions aligned with the input fabric.



Fig. 15. Additional qualitative results of our method. The foreground mask denotes the garment region in the input image, and the fabric swatch in the lower-left corner represents the target fabric for try-on. Our method robustly handles a wide range of garment shapes and deformations, lighting and illumination conditions, and diverse texture patterns. Image credits (via Unsplash): Row 1 Column 1 © Korie Cull; Row 1, Column 4 © Ruan Richard Rodrigues; Row 2, Column 1 © Colton Sturgeon; Row 2, Column 4 © Luigi Ritchie; Row 3, Column 1 © Leela Shyam; Row 3, Column 4 © Zayed Ahmed Zadu; Row 4, Column 1 © Ali Choubin; and Row 4, Column 4 © Dynamic Wang.




Fig. 16. Additional qualitative results of our method.



Fig. 17. Additional qualitative results of our method.




Fig. 18. Additional Results of Tile Augmentation. Our tile augmentation module generates seamless fabric tiles that closely match the input fabric, preserving high-frequency textures and woven structures.

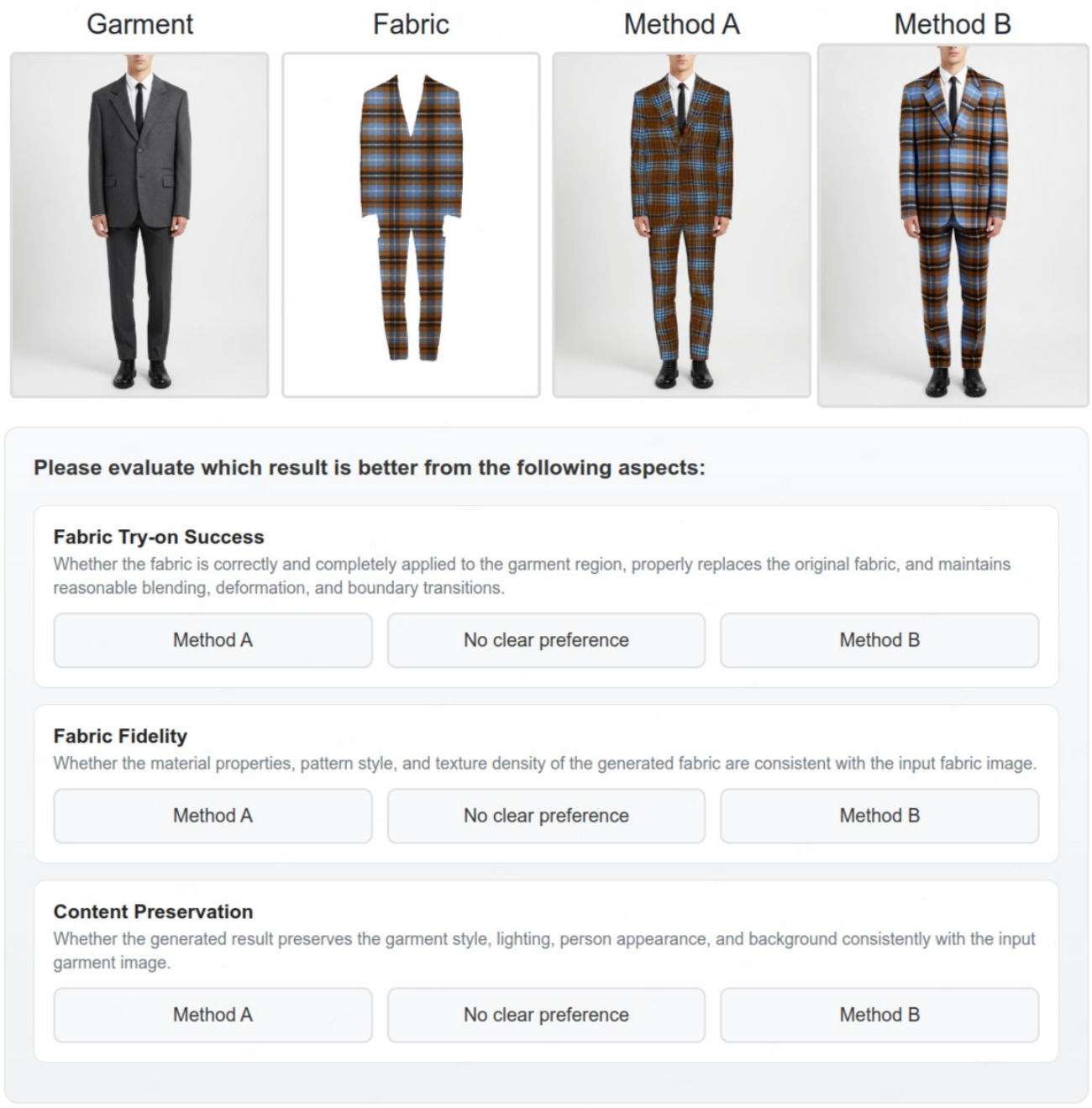


Fig. 19. User study interface.