NI-Tex: Non-isometric Image-based Garment Texture Generation

Supplementary Material

A. Implementation Details

A.1. Dataset Curation Details

We use the Nano Banana to generate about **50K** edited images for frames sampled in 3D garment videos. We use a VLM (Qwen3-VL) to quality-check Nano Banana-generated image edits based on our predefined principles. This enables automatic removal of obviously incorrect edits, followed by a manual process to refine the selections.

A.2. MR Rectification for Cross-pose Supervision

Since the metallic and roughness values vary across frames, their reflective behaviors are inconsistent, making direct supervision from the supervision 3D frame unreliable. To address this, we introduce MR rectification (Figure 1). We sample a representative foreground pixel from the MR image of the condition 3D frame and replace all foreground regions of the supervision frame's MR images with its value, enabling consistent cross-frame supervision during training.

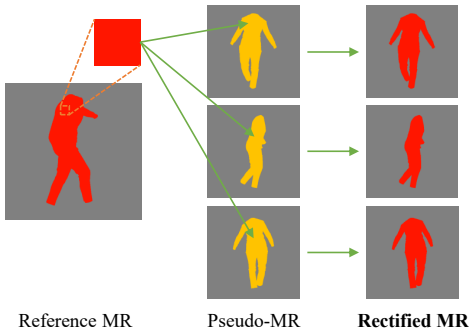

Figure 1. We randomly select one MR image from the condition 3D frame as the reference MR image. Since each MR map is assigned a globally uniform value, we can extract any foreground pixel as the reference pixel. We then index the MR images of the supervision 3D frame and replace all foreground pixels with the reference value, enabling cross-frame supervision.

A.3. Iterative Baking Algorithm Details

In this subsection, we describe the implementation details of the baking algorithm (see Figure 2). Specifically, we talk about the training of UQ model, the details of the reweighting of multi-view texture maps.

UQ Model Training. We adopt the resnet-50 architecture that predicts per-pixel uncertainty from an single-view rendered texture map. The training dataset is constructed from pairs of rendered texture map (intermediate baking result from the baking algorithm) and the ground truth texture map of the same view. The data pairs come from an an error simulation procedure describe in section 4.2. To quantify the difference between the rendered texture map and the ground truth texture map, we use the SSIM metric to quantify the per-pixel uncertainty map from two texture maps. Specifically, the compute the per-pixel SSIM values are in the range $[0, 1]$. The supervision loss is simply:

$$\sum_{p_i} \|\mathrm{UQ}(p_i) - y^{\mathrm{SSIM,\,GT}}(p_i)\|_2^2 \tag{1}$$

where p_i is a pixel, $\mathrm{UQ}(p_i)$ denotes the predicted uncertainty value and $y^{\mathrm{SSIM,\,GT}}(p_i)$ is the SSIM value computed by comparing with the ground truth texture map.

Multi-view Reweighting. The UQ model is also used to reweight texture maps based on the predicted uncertainty scores. When performing multi-view blending, we compute the final texture $t_i^\star$ from the texture map of multiple views $\{p_{ij}|j\}$, where each view j is weighted by the uncertainty score $(1 - \mathrm{UQ}(p_{ij}))$ and a constant view score c_j corresponding to how far the view is from the frontal and back view. For the front and back viewpoints, we set $c_j = 1$. For other viewpoints, c_j is progressively attenuated to 0.5, 0.25, 0.125, and 0.1 according to their relative distance and perceptual importance.

$$t_i^\star = \frac{\sum_j (1 - \mathrm{UQ}(p_{ij}))c_j p_{ij}}{\sum_j (1 - \mathrm{UQ}(p_{ij}))c_j + \epsilon_1} \tag{2}$$

A.4. Training Data Preparation Details

We encode roughness and metallic into an RGB image, where the R channel is fixed to 255, and the G and B channels store the corresponding values scaled by 255. For each frame, we randomly apply different types of lighting sources, including point lights, area lights, and environment lights. We render the garment under 10 views, capturing the illumination effects together with material-related attributes (e.g., PBR texture properties) and geometry-related attributes (e.g., normal images and position images).

Since the MR values in our 3D garment videos are globally uniform, the model tends to overfit these constant properties during training. To enhance the network's perception and generalization across different MR materials, we incorporate additional supervision from the Objaverse and TexVerse datasets for cross-mixed training.

We have ever carefully considered whether to use complex real-world datasets to expand our dataset. However,

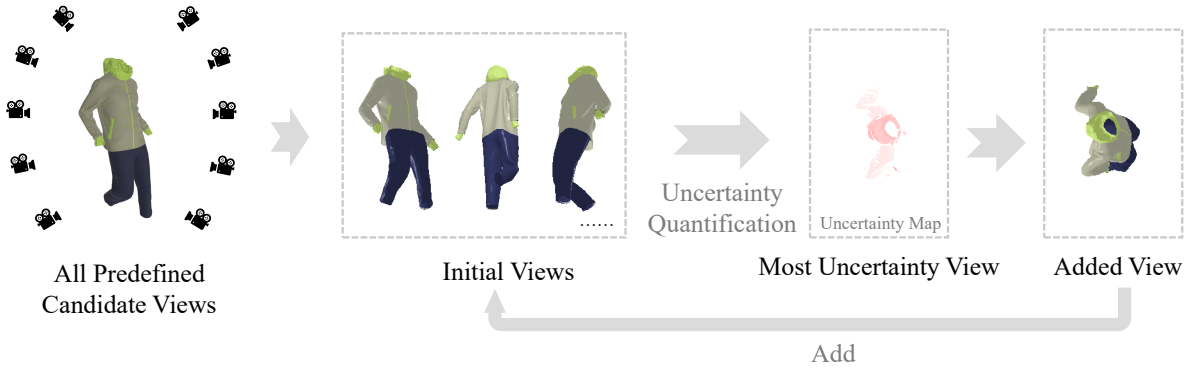

Figure 2. Given a set of predefined candidate views, our view selection algorithm computes an overall uncertainty score for each view by averaging its per-pixel uncertainty. It then selects the view that exhibits the highest uncertainty.

the GT albedo and MR of real-world datasets are difficult to obtain, as these attributes are typically applied in graphics **simulators**. We tried to use 2D generative methods to estimate their albedo and MR as GT, which proved to be unstable. Therefore, we ultimately decided not to use real datasets during training.

B. Experiments

B.1. Generation for Industrial Meshes

In Figure 5, we supplement additional examples using wild images from DeepFashion2 as image prompts, with industrial meshes as the target meshes. We find that NI-Tex is capable of reliably generating textures that closely conform to the input image prompts, maintaining high fidelity even in highly challenging conditions.

B.2. Generation for Generated Meshes

In Figure 6, we present additional examples using wild images from DeepFashion2 as image prompts, with Hunyuan-generated meshes as the target geometry. We observe that NI-Tex effectively preserves logos and local details, and can also faithfully maintain complex patterns such as spots.

B.3. Multi-view Visualization

To further verify the 3D consistency of the textures generated by NI-Tex, we project the texture maps back onto the mesh surface. In addition to the front and back views, we further show four additional viewpoints: front-left, back-left, front-right, and back-right. Results in this subsection are displayed under lighting conditions and some of the examples are taken from previous experiments. (See Figure7, 8, 9)

B.4. Baking Strategy

After training the UQ model, we compare our view selection strategy with the coverage-based view selection baseline. We separately use the two greedy metrics, i^{UQ} and i^{cvg}, to select 10 views on the given Bedlam geometry and bake. We select the worst viewpoints as our test views and perform quantitative evaluation using PSNR. Figure 4 shows the 10 selected views for each strategy, along with their corresponding final test views and PSNR values.

Figure 3. **Failure cases.** Extreme image prompt deformations lead to synthesis artifacts.

B.5. Failure Case Analysis

The generated multi-view images occasionally suffer from artifacts when conditioned on severely deformed image prompts (see Figure 3). We hypothesize that scaling up the volume and diversity of the training data will significantly improve the synthesis quality of the NI-Tex multi-view results.

C. Limitations and Future work.

While NI-Tex delivers high-quality non-isometric texture generation, its generalization to complex rigid deformations remains limited due to the lack of physically simulated data for general objects. In future work, we aim to enhance the model's 3D self-awareness of object deformation, enabling

Figure 4. We compare the 10 views selected by the coverage-based strategy with those selected by our strategy. We then choose the worst viewpoint as the test view and compute its PSNR.

more robust non-isometric texture generation under limited data.

Figure 5. Texture generation results on industrial meshes using wild images from DeepFashion2 as image prompts. NI-Tex effectively captures the appearance of the input images, even under challenging variations. ('/' indicates generation failure.)

Figure 6. Texture generation results on Hunyuan-generated meshes using wild images from DeepFashion2 as image prompts. NI-Tex demonstrates strong capability in retaining fine-grained details, including logos and intricate patterns, while accurately capturing textures across diverse clothing types.

Figure 7. Multi-view visualization for industrial meshes (using well-render image prompts).

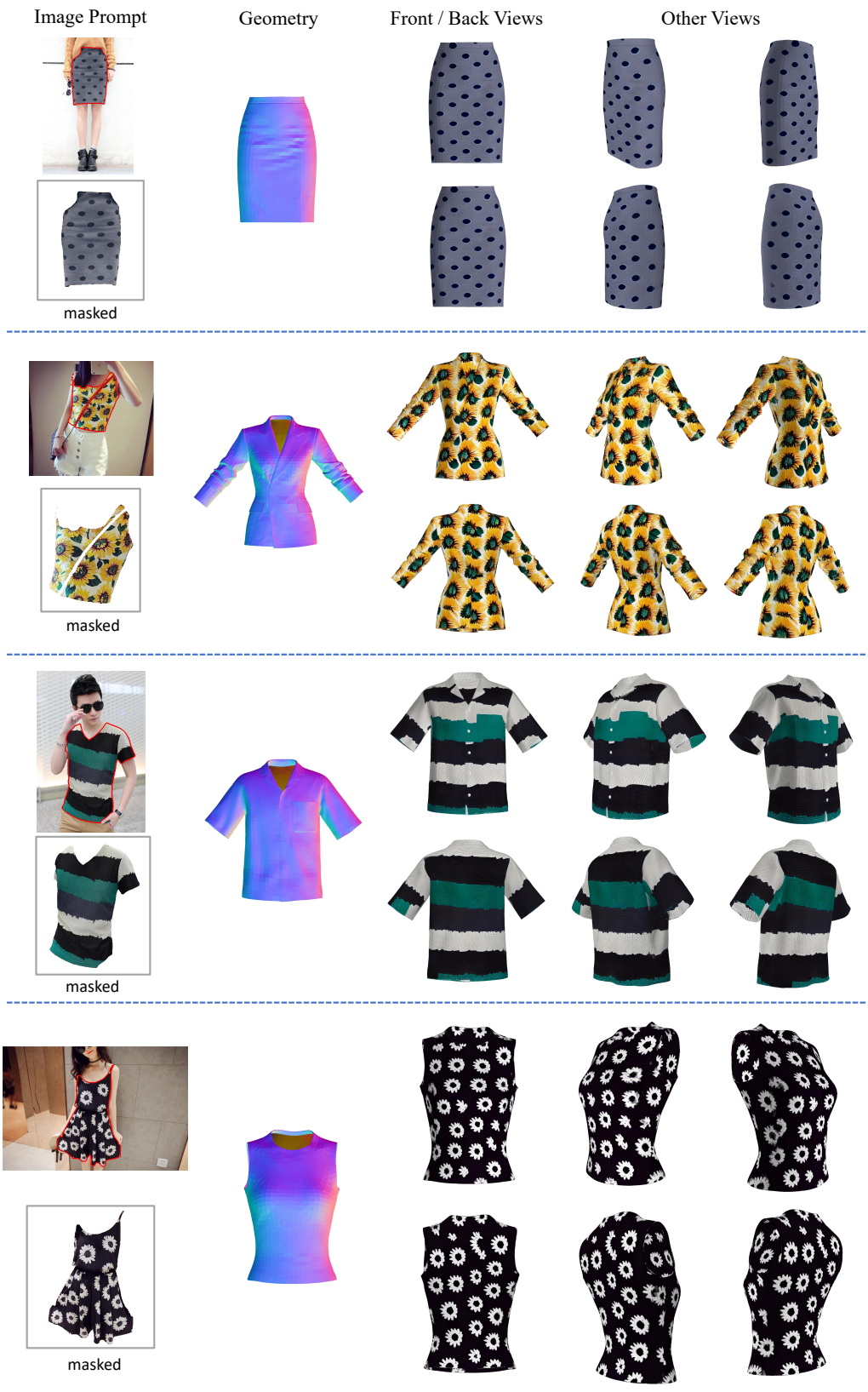

Figure 8. Multi-view visualization for industrial meshes (using image prompts from DeepFashion2).

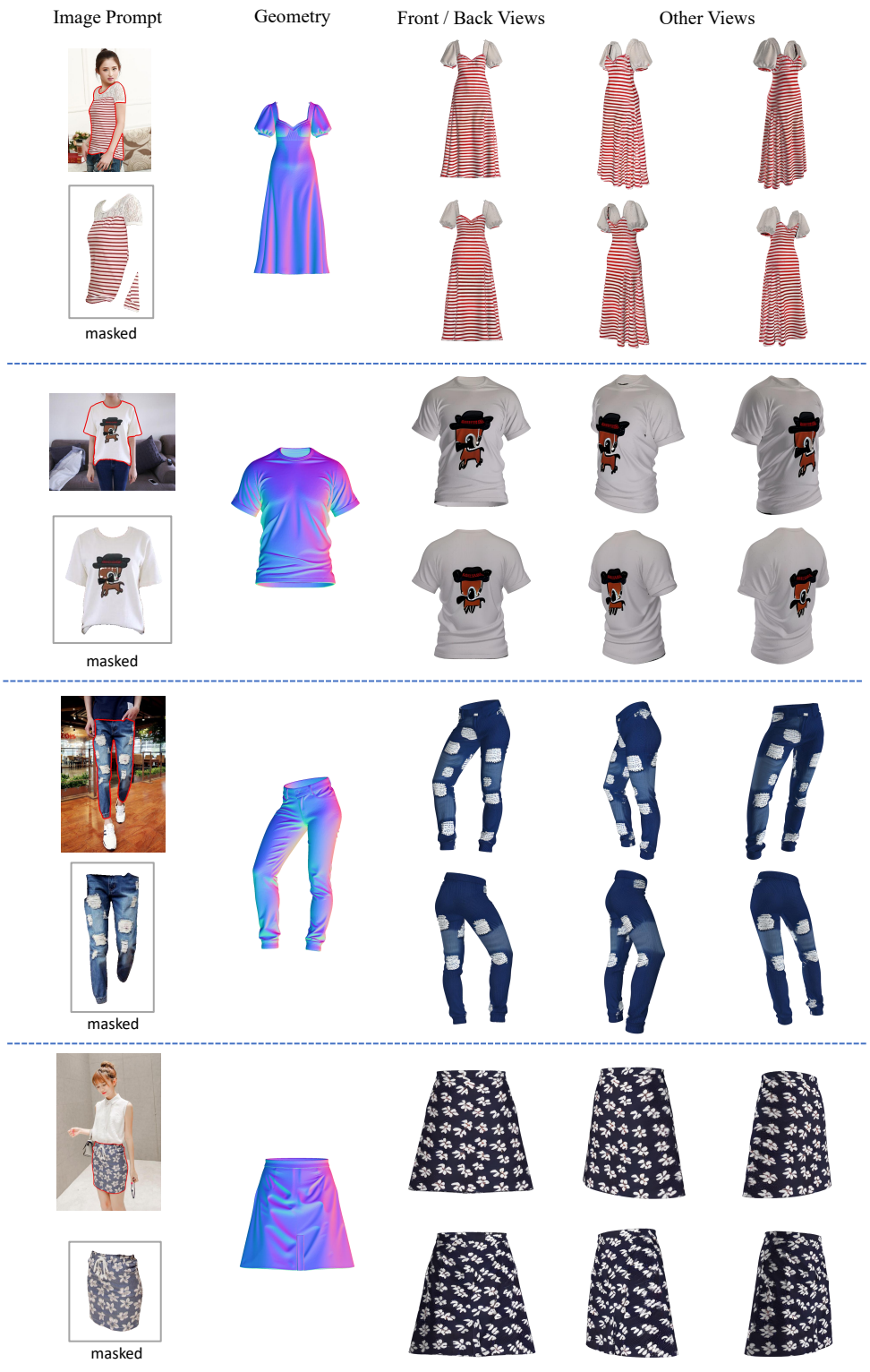

Figure 9. Multi-view visualization for Hunyuan-generated meshes (using image prompts from DeepFashion2).