Shining Yourself: High-Fidelity Ornaments Virtual Try-on with Diffusion Model

Supplementary Material

6. Additional Experiments

6.1. Comparison on Necklaces, Earrings, and Rings

In addition to the comparison with previous methods using the bracelet dataset, as discussed in the main text, we extended the evaluation to other categories of ornaments. The results are shown in Fig. 9. Consistent with the qualitative comparisons in the main text, Paint-by-Example [34] faces challenges in preserving geometric or ID information, retaining only partial semantic features. AnyDoor [6] has difficulty capturing correct wearing patterns and natural region inpainting. IDM-VTON [7] shows some improvement over other methods but still struggles with maintaining fine details and spatial relationships.

In contrast, our method demonstrates superior performance across all categories. Necklaces, due to their small wearing area and the challenge of invisible chain parts, are the most difficult among all categories. Nevertheless, our method achieves good results. Notably, in the first row of examples, where some structures are not visible in the reference image, our method faithfully preserves the reference ornament's ID information. Rings share a similar wearing pattern to bracelets but include more subtle structural details. As seen in the 5th and 6th rows, our method generates high-fidelity try-on images that preserve these fine details. For earrings, the results in the 8th row demonstrate that our method effectively handles fine linear decorations.

6.2. Additional Results

We conducted further experiments across all categories, and Fig. 10 presents the results of our virtual try-on. Our method demonstrates stable, high-fidelity virtual try-on for ornaments. Additionally, we explored cross-domain ornament virtual try-on and found that our method is capable of virtual try-on for certain animated characters. This showcases the robustness of our approach and indicates that, with sufficient data, our method can be extended to images from other domains.

6.3. Additional Ablation Study

Experiments to Validate Our Motivation. To validate the motivation for mask refinement, we conducted experiments on different masks using the baseline based on ReferenceNet and Stable Diffusion. The results are shown in Fig. 7. We trained the model with various input masks, including bounding boxes, oriented bounding boxes (OBB), convex hulls, and ground truth masks. The results show that as mask refinement increases, the fidelity and ID consistency of the generated images improve. For instance, using a bounding box as the mask increases the probability of errors in the wearing pattern. The OBB, which adds limited pose information, shows some improvement but still fails to produce satisfactory results. The convex hull adds extra shape information, and some geometric structures, which earlier methods struggled to preserve, are retained in the output. The ground truth mask yields the best results, providing precise shape and location information. These results highlight the importance of pose and accurate masks for generation quality, which is a key idea in our work. However, as mentioned in Section 3.2 of the main text, a fine-grained wearing mask cannot be obtained during the inference process. Our method uses only the bounding box as input to predict the wearing mask and then refines the original mask. The results demonstrate that our method achieves performance close to that of the ground truth mask.

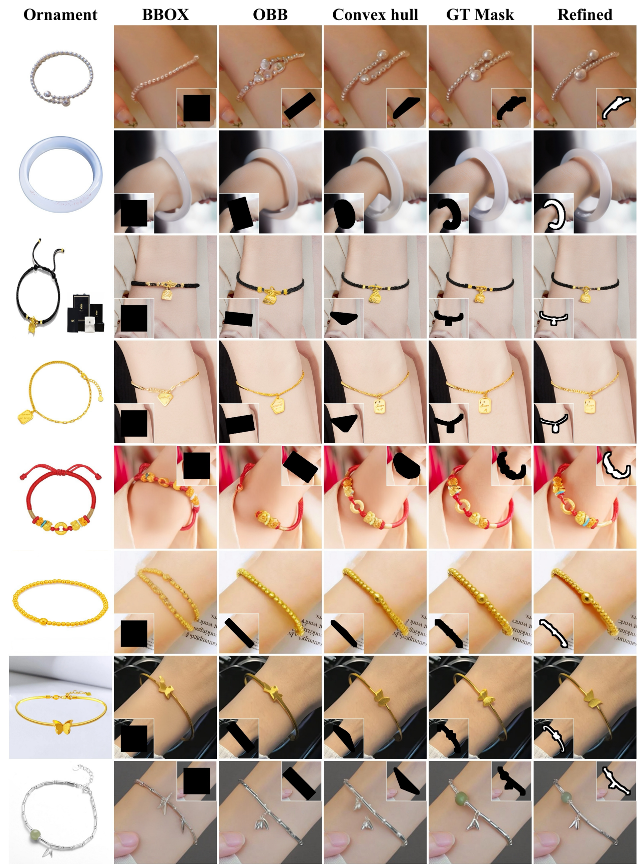

Figure 7. Experiments to validate our motivation.

6.4. Results on Garments Virtual Try-on

Our task extends the field of garments virtual try-on. As analyzed in the introduction, the proposed ornaments virtual try-on presents more significant challenges compared to garments virtual try-on. Therefore, we have developed a series of customized improvement strategies. To demonstrate the performance of our method in garments virtual try-on, we conducted qualitative experiments on the VITON-HD dataset, and the comparative results with IDM-VTON are shown in Fig.8.

It should be noted that our method is not specifically designed for garments virtual try-on, and thus many inputs for garments virtual try-on are not considered. We only utilize the model image, garments image, and mask. Despite this limitation, our method achieves comparable results to state-of-the-art (SOTA) methods. Given that garments virtual try-on has been extensively studied and previous methods have achieved remarkable results, our method faces certain limitations in further enhancing garments try-on performance.

Through multiple experimental validations presented in this paper, we can conclude that our method exhibits excellent versatility. It maintains the performance of garments virtual try-on while achieving outstanding results in ornaments virtual try-on. This characteristic makes our method highly promising for a wide range of applications in the field of virtual try-on.

IDM-VTON

Ours

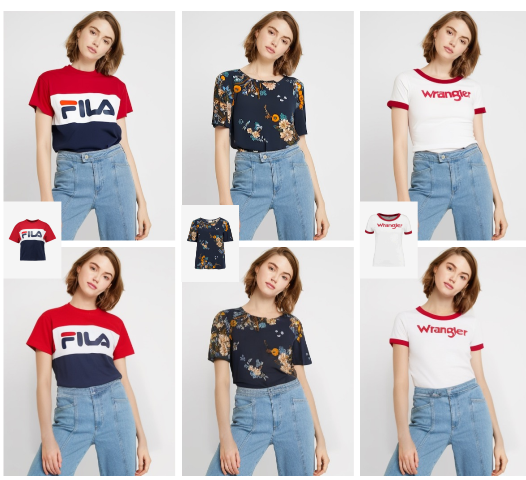

Figure 8. Visual comparison on garments virtual try-on.

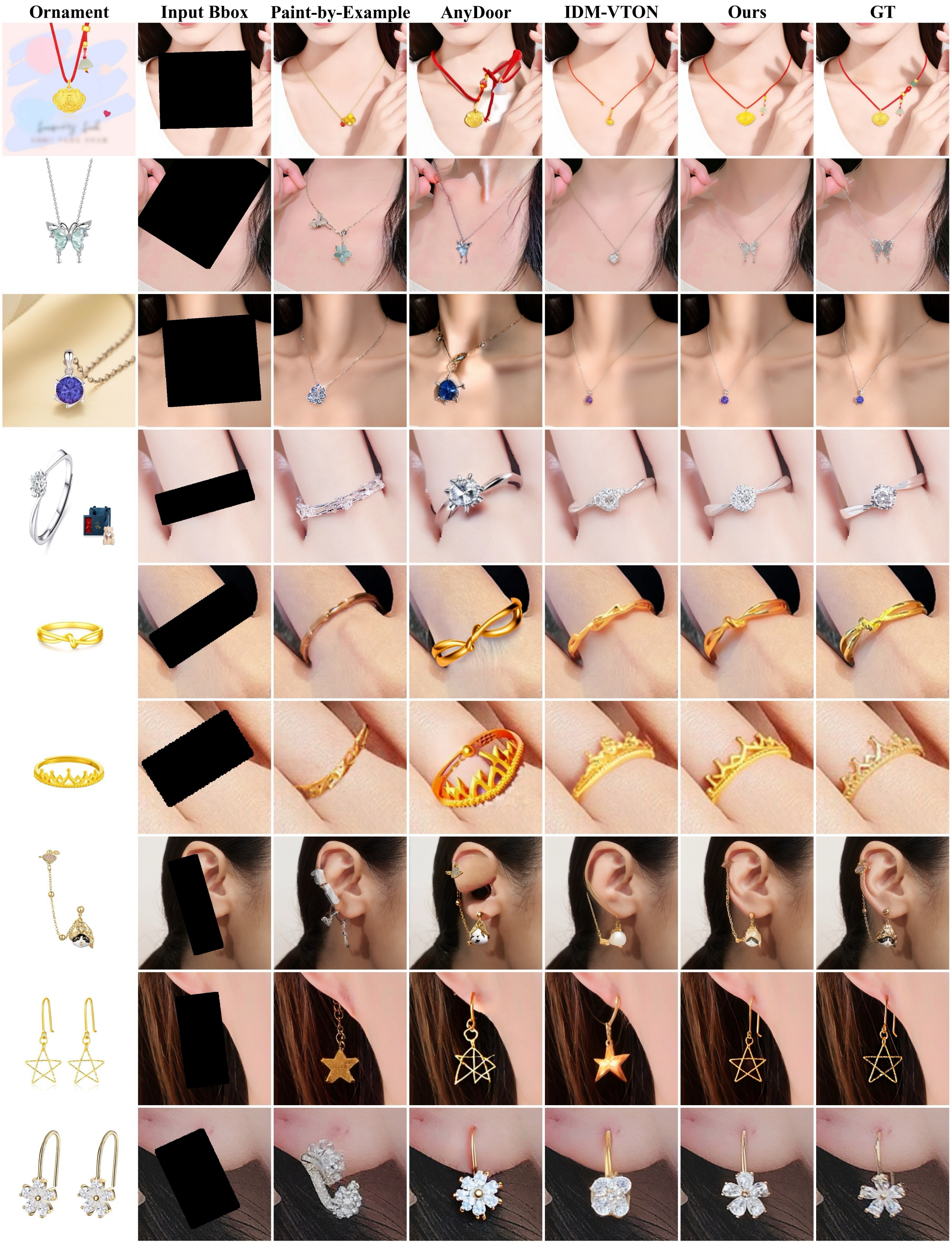

Figure 9. Visual comparison on necklaces, rings, and earrings.

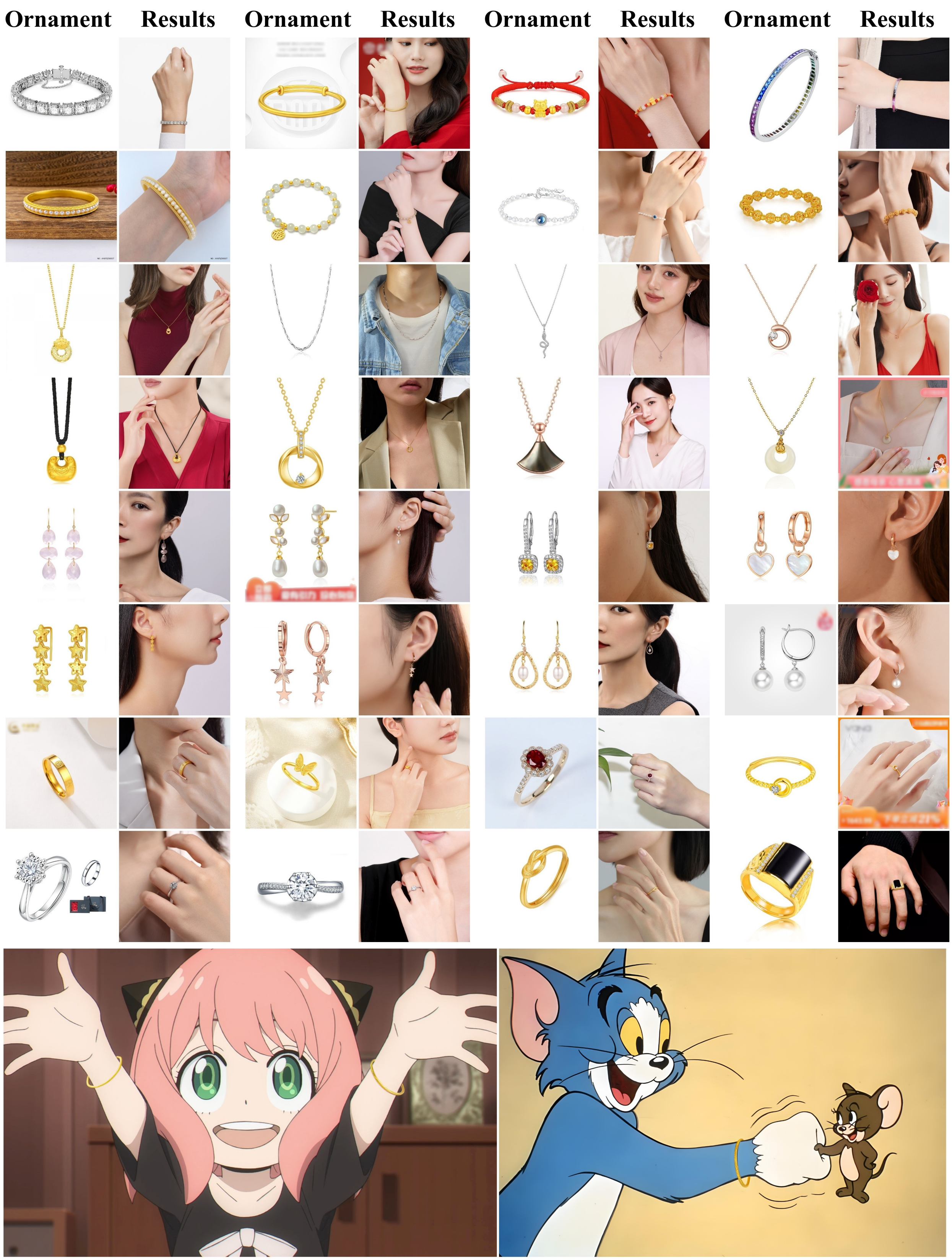

Figure 10. More results.