Supplementary Material

In this supplementary material, we provide additional visual examples and discussions to further illustrate the effectiveness of our proposed HandCraft framework. We start by presenting more visual comparison results of malformed and restored hands across various styles and scenarios. Next, we explore several challenging cases, including multiple hands of a single person, multiple people with multiple hands, and hands with occlusions, highlighting the current limitations of our approach and providing direction for future research efforts. Finally, we showcase instances of anatomically incorrect hands produced by the latest state-of-the-art generative models, such as Stable Diffusion XL [24] and Sora [6], emphasizing the persistent nature of the malformed hand problem and the continued relevance of HandCraft in addressing this issue. Through these supplementary materials, we aim to provide a more comprehensive understanding of HandCraft's capabilities, limitations, and potential future developments.

A. Visual Comparison of Malformed and Restored Hands

To further demonstrate the effectiveness of HandCraft in restoring malformed hands, we present a visual comparison of images before and after applying our method. Fig. 9 showcases a diverse set of examples from the MalHand-realistic and MalHand-artistic datasets, covering various styles and hand malformations. These visual examples demonstrate HandCraft's ability to handle a wide range of hand malformations across different artistic styles. The restored hands exhibit improved anatomical correctness and seamless integration with the original image, preserving the overall aesthetic and maintaining consistency with the intended style.

B. Challenging Cases Including Failure Cases

While HandCraft demonstrates strong performance in restoring malformed hands in a variety of scenarios, there remain several challenging cases that worth further discussion and investigation. These cases highlight the current limitations of our approach and provide direction for future research efforts in this domain.

B.1. Multiple Hands of a Single Person

One challenging scenario arises when a single person has multiple hands visible in the image, as shown in Fig. 10. We observe that the success of hand restoration in such cases often depends on the accuracy of the hand pose estimation. When the hands are not overlapping or intertwined, the hand pose estimation algorithm can typically provide correct hand keypoints, leading to successful restoration. However, when the hands are in close proximity or crossing over each other (i.e., interhand occlusion), hand pose estimation becomes more challenging, and the keypoints may not be accurately estimated. Consequently, the restoration quality for interhand scenarios is generally lower compared to non-overlapping hands.

B.2. Multiple People with Multiple Hands

Another challenging case involves images depicting multiple people, each with multiple visible hands, as presented in Fig. 11. Similar to the single-person scenario, the restoration results in such cases can be influenced by the accuracy of the hand pose estimation. When hands from different individuals overlap or intersect, the restoration quality tends to suffer. The complexity introduced by the presence of multiple hands belonging to different individuals poses additional difficulties for accurate hand pose estimation and subsequent restoration.

B.3. Hands with Occlusions

Hand occlusion presents another significant challenge for our hand restoration approach, as illustrated in Fig. 12. When parts of the hand are obscured by other objects in the image, the available information for hand pose estimation and restoration becomes limited. The absence of complete hand visibility can lead to inaccurate keypoint estimations and suboptimal restoration results. Handling hand occlusions effectively requires more advanced techniques for inferring the complete hand structure from partial observations. In some cases, the occluding object may disappear or be partially removed as a consequence of the restoration process, as the model focuses on reconstructing the hand without considering the context of the occluding object.

B.4. Persistent Malformed Hands in Generative Models

The rapid advancements in generative image models have led to remarkable improvements in the quality and realism of AI-generated imagery. However, despite the progress made, the issue of malformed hands has remained a persistent challenge across various generations of these models. Even the recently-released state-of-the-art systems, such as Stable Diffusion XL [24] and Sora [6], are not immune to this problem.

In the supplementary materials, we provide visualizations that highlight instances of anatomically incorrect hands produced by these latest models. For example, portraits generated using SDXL [24] exhibit hands with abnormalities and malformations, such as pinky fingers that are noticeably shorter than normal, hands with six fingers, and unnatural connections between fingers that resemble fusion or webbing. Similarly, images generated by Sora [6] also display hand malformations, including fingers bent back-

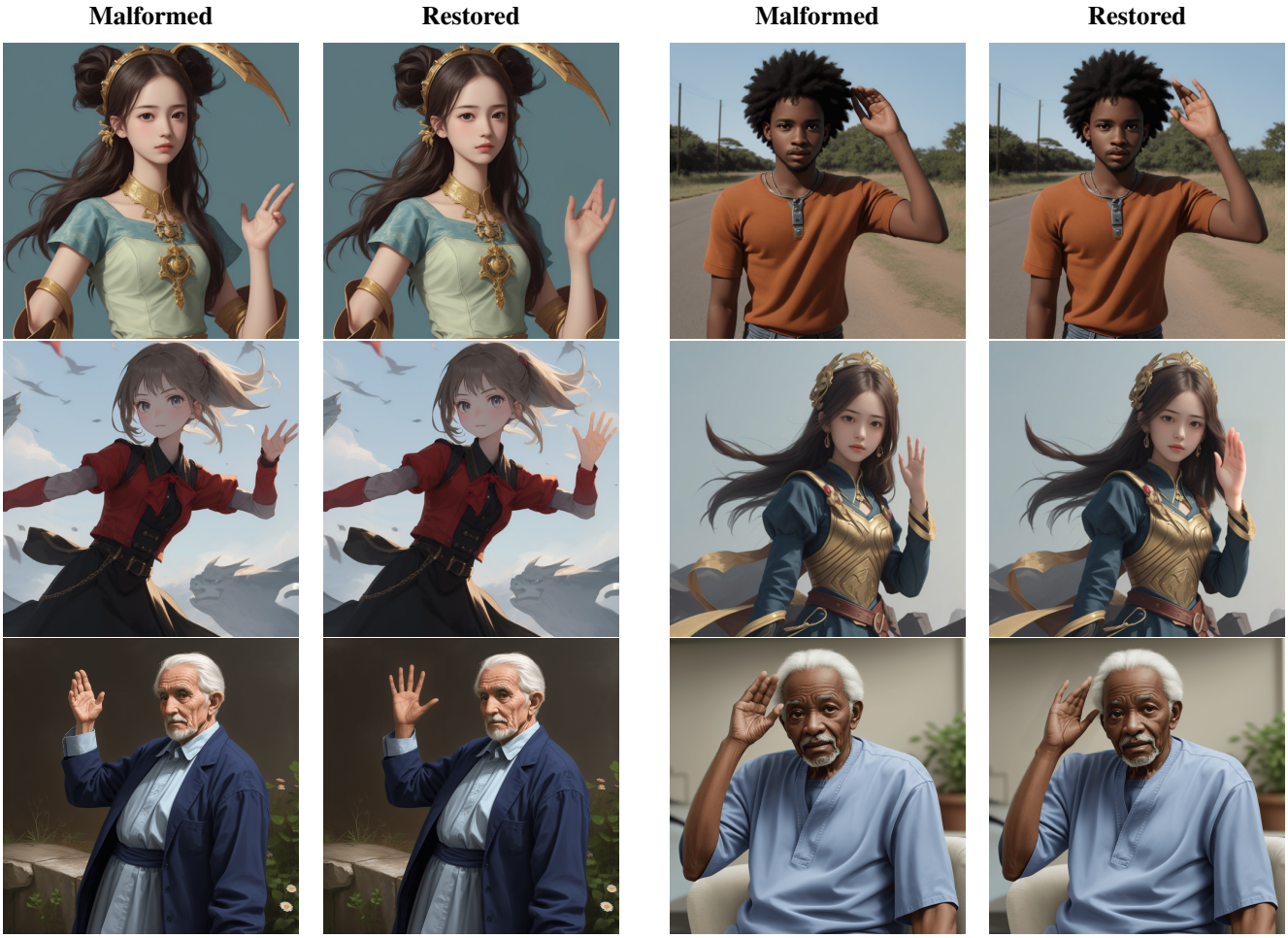


Figure 9. **Visual comparison of malformed and restored hands using HandCraft.** The left column shows images with malformed hands from the MalHand-realistic and MalHand-artistic datasets, while the right column presents the corresponding restored images. HandCraft successfully corrects various types of hand malformations, such as abnormal finger lengths, extra fingers, and anatomically impossible configurations, resulting in more realistic and aesthetically pleasing hands that integrate seamlessly with the original image style.

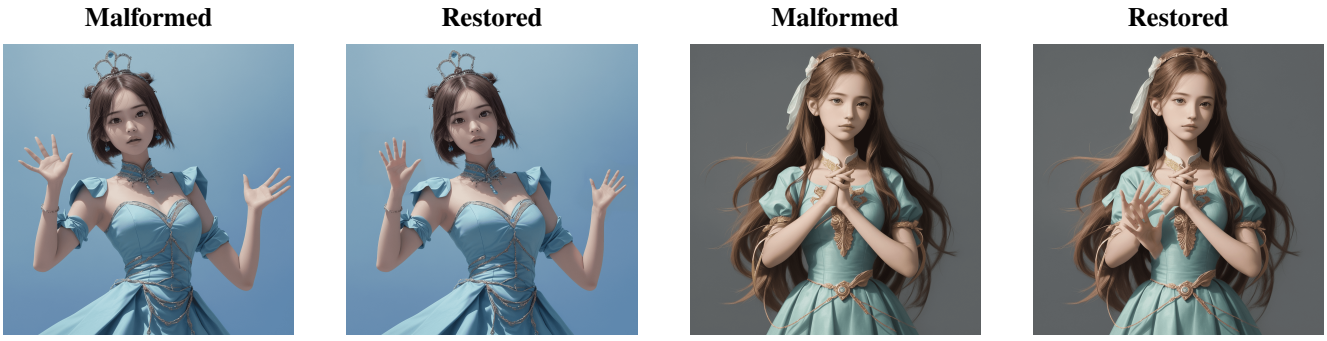


Figure 10. **Hand Restoration with Multiple Hands of a Single Person.** The first pair shows a case where the person's hands are not overlapping, leading to successful restoration. The second pair presents a case where the hands are in close proximity, resulting in a more challenging restoration scenario due to potential interhand occlusion and inaccurate hand pose estimation.

wards at impossible angles, abnormally long thumbs disproportionate to the rest of the hand, and hands with two thumbs.

These examples underscore the pervasive nature of the

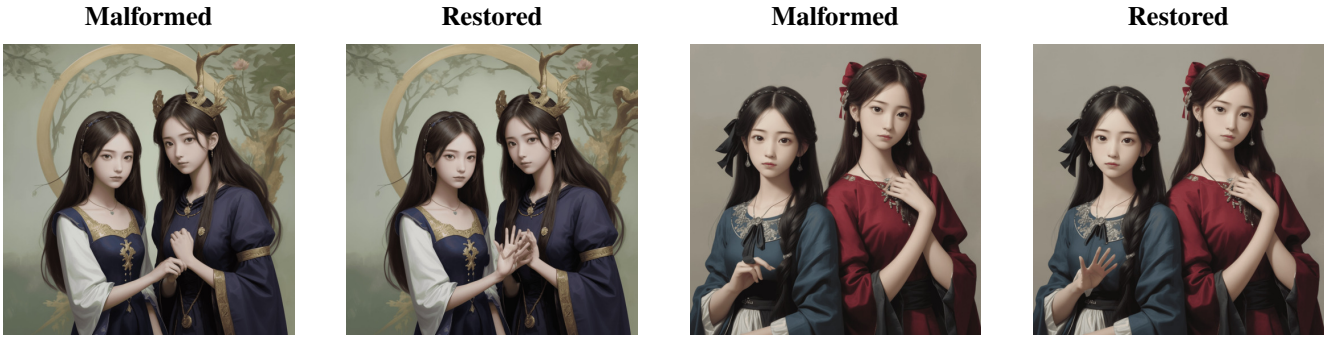


Figure 11. **Hand Restoration with Multiple People and Multiple Hands.** The images demonstrate the challenge of restoring hands in scenarios involving multiple individuals, each with multiple visible hands. The restoration quality can be affected when hands from different people overlap or intersect, as seen in the second pair of images. Accurate hand pose estimation becomes more difficult in such cases, leading to potential limitations in the restoration process.

Figure 12. **Hand Restoration with Occlusion.** The first pair shows a case where the hand is partially occluded. Despite this occlusion, HandCraft is able to generate a plausible restored hand that aligns with the visible portion of the hand and maintains a natural appearance. The second pair presents a case where the hand is interacting with an object (holding a glass). HandCraft successfully restores the hand in this scenario, yet unable to maintain the original object.

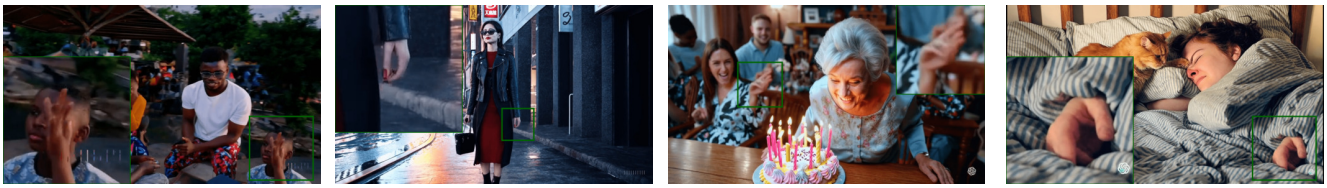


Figure 13. **Examples of images generated by Sora [6] that exhibit hand malformations.** (a) Fingers bent backwards at an impossible angle towards the back of the hand. (b) Abnormally long thumb disproportionate to the rest of the hand. (c) Two thumbs present on the same hand. (d) Unnaturally thick connection between the fingers and palm, deviating from normal human hand proportions.

malformed hand problem, which has persisted despite the advancements in generative image models. The complexity of hand anatomy and the fine-grained details required to accurately depict hands pose significant challenges for these models. As a result, we expect HandCraft to remain a useful tool for the foreseeable future, as the issue of malformed hands appears to be quite persistent across generations of generative models.

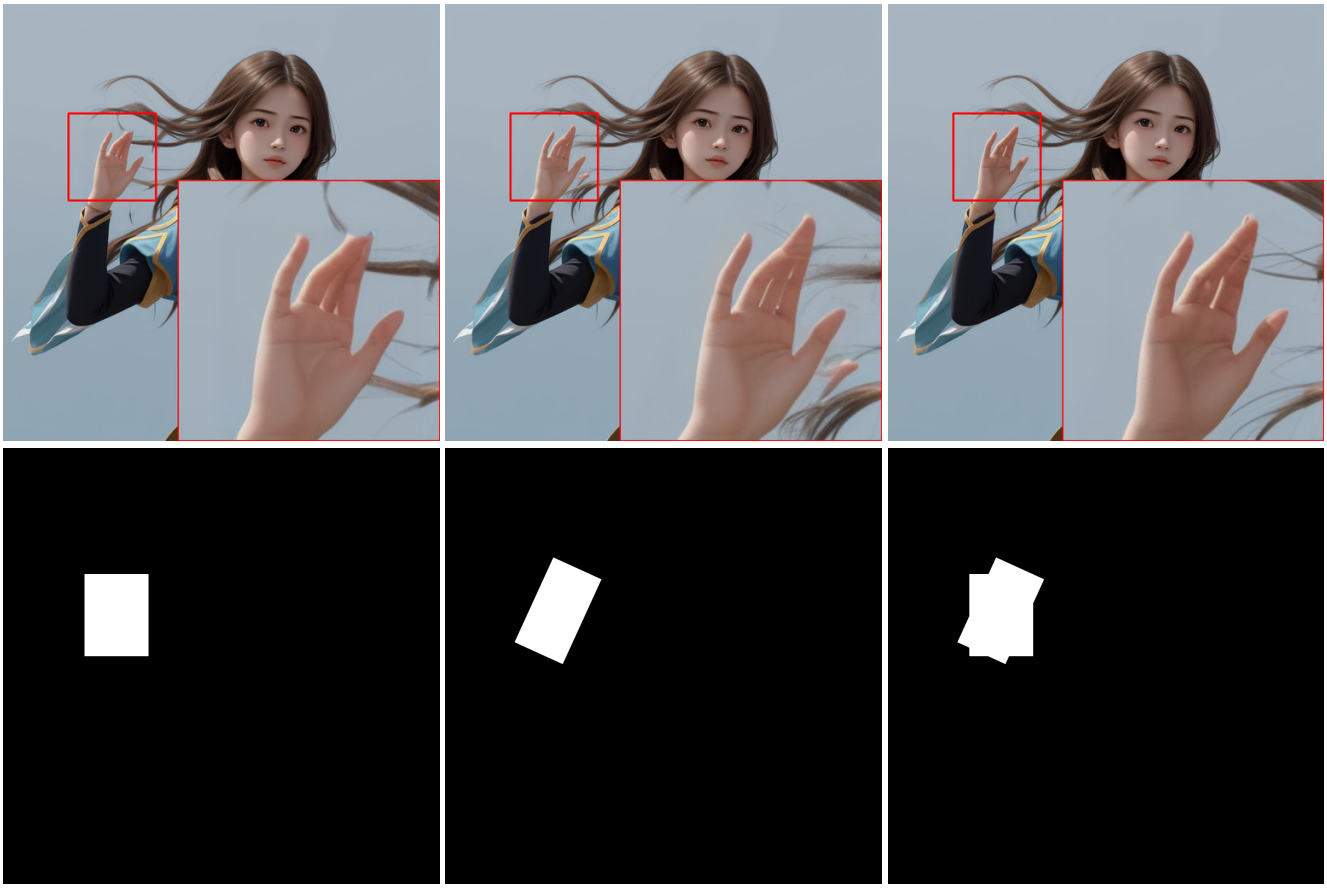
Moreover, even if future generative models overcome the challenge of producing anatomically correct hands consistently, HandCraft offers functionality that extends beyond simply correcting malformed hands. Our method provides users with the ability to change hand gestures and poses for creative control, allowing them to customize and refine the generated images according to their specific needs and preferences. This added flexibility makes HandCraft a valuable tool for artists, designers, and content creators who seek to enhance the expressiveness and versatility of AI-generated imagery involving human hand articulation.

In conclusion, the persistent nature of the malformed

(a) (b) (c) (d)

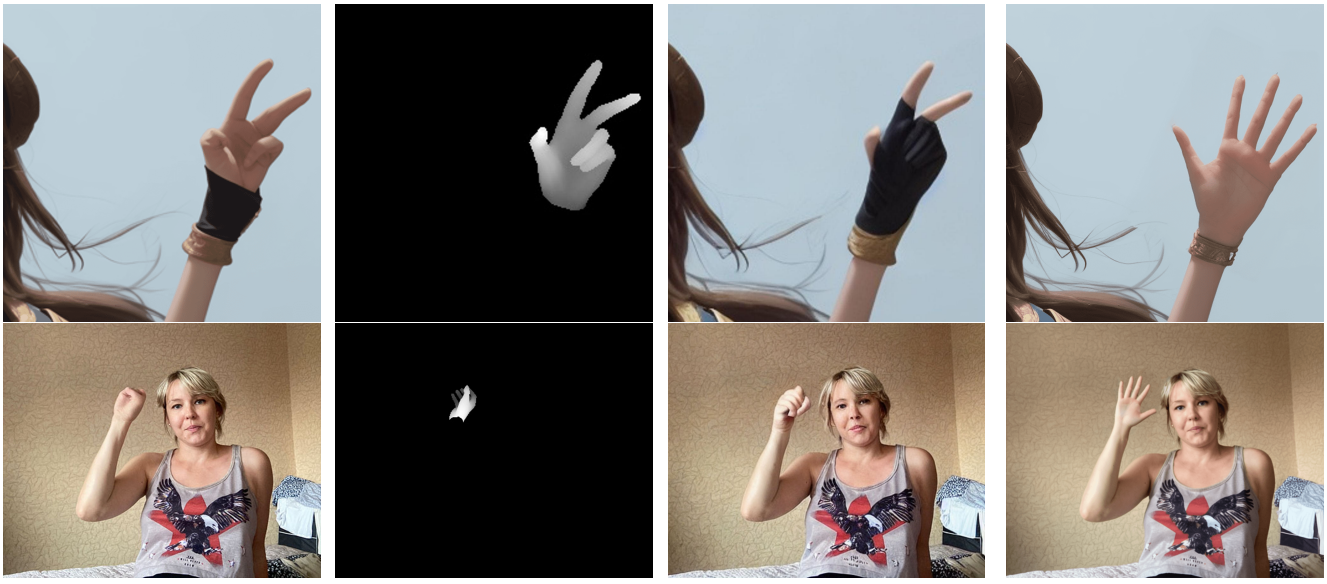
Figure 14. **Portraits generated using SDXL [24], and the hands in these images exhibit anatomical abnormalities and malformations.** (a) The pinky finger on the left hand is noticeably shorter than it should be for a normal hand. (b) The hand has 6 fingers. (c) This hand also has 6 fingers, but one of the fingers appears to be partially missing or disappeared, leaving only a vestigial remnant. (d) The pinky finger is also significantly shorter than normal; additionally, the connection between the index and middle fingers looks unnatural, as if there is some fusion or webbing present.

hand problem in generative image models highlights the importance and relevance of HandCraft. By addressing this issue and providing a simple yet effective solution, our framework contributes to the ongoing efforts to improve the realism and controllability of AI-generated imagery. Furthermore, the additional functionality of allowing users to manipulate hand gestures and poses positions HandCraft as a valuable tool for creative applications, even as generative models continue to evolve and advance.



(a) Only the bounding box mask (b) Only the control mask (c) Union

Figure 15. Ablation study on the effect of using different masks for hand restoration. (a) Using only the bounding box mask leads to missing fingertips in the restored hand. (b) Using only the control mask retains parts of the original malformed hand. (c) Using the union of the bounding box and control masks achieves the most natural and complete hand restoration. The top row shows the corresponding restored hand images, while the bottom row displays the different masks used (red box indicates the detection bounding box, beige overlay shows the control mask).



(a) Malformed hand (b) Mesh fitting (c) Restore via mesh (d) Using template

Figure 16. Comparison of hand restoration methods: (a) The original image with a deformed hand where fingers are bent in an unnatural manner or are missing. (b) The result of mesh fitting, mimicking the incorrect finger alignment and positioning from the original, resulting in a hand orientation that does not match the natural pose. (c) The outcome of attempting restoration with the flawed mesh, maintaining the unnatural bending of the fingers, or resulting in a malformed hand inconsistent with the mesh condition. (d) The hand restored using a predefined template, which achieves a natural-looking hand pose and maintains anatomical accuracy.