Supplementary Materials for InstructGIE

Zichong Meng*[1], Changdi Yang*[1], Jun Liu[1], Hao Tang**[2], Pu Zhao**[1], and Yanzhi Wang**[1]

[1] Northeastern University, Boston MA 02115, USA
[2] Carnegie Mellon University, Pittsburgh PA 15213, USA

We further present our proposed method by providing the following supplemental materials:

- Appendix A: Ablation Study of Ways to Mask for Visual Prompted Conditions
- Appendix B: Edges Cases and Results
- Appendix C: Failure Cases and Potential Future Improvement
- Appendix D: Additional Image Editing Result Visualization

A Ablation Study of Ways to Mask for Visual Prompted Conditions

In this section, we present ablation studies on different ways for visual prompt masking during training. We show different masking methods and their qualitative results in Fig. A1. In this ablation study, we show 5 different masking strategies: White Mask, Random Noise Mask, Grey Mask (ours), Random Patch Mask and Black Mask. Since the mask is concatenated with example pair image and query image together as one input to the Stable Diffusion model, the value distribution of pixels in the mask will affect the value distribution of the output. In Fig. A1, we can observe that white and black masks will make the output images brighter and darker, respectively. The random noise mask makes the output images look more distorted. The random patch mask results in generation of random part of query image. Overall, the grey mask achieves the best quality.

B Edges Cases and Results

InstructGIE adopts both visual and text prompts to elevate the image editing quality with boosted in-context learning capability, unified language instructions, and selective area matching. In Fig. A2, we show InstructGIE's capability on edge cases including out-of-domain images with real-world photo pairs of various qualities/conditions.

---

* Equal Contribution
** Corresponding Authors

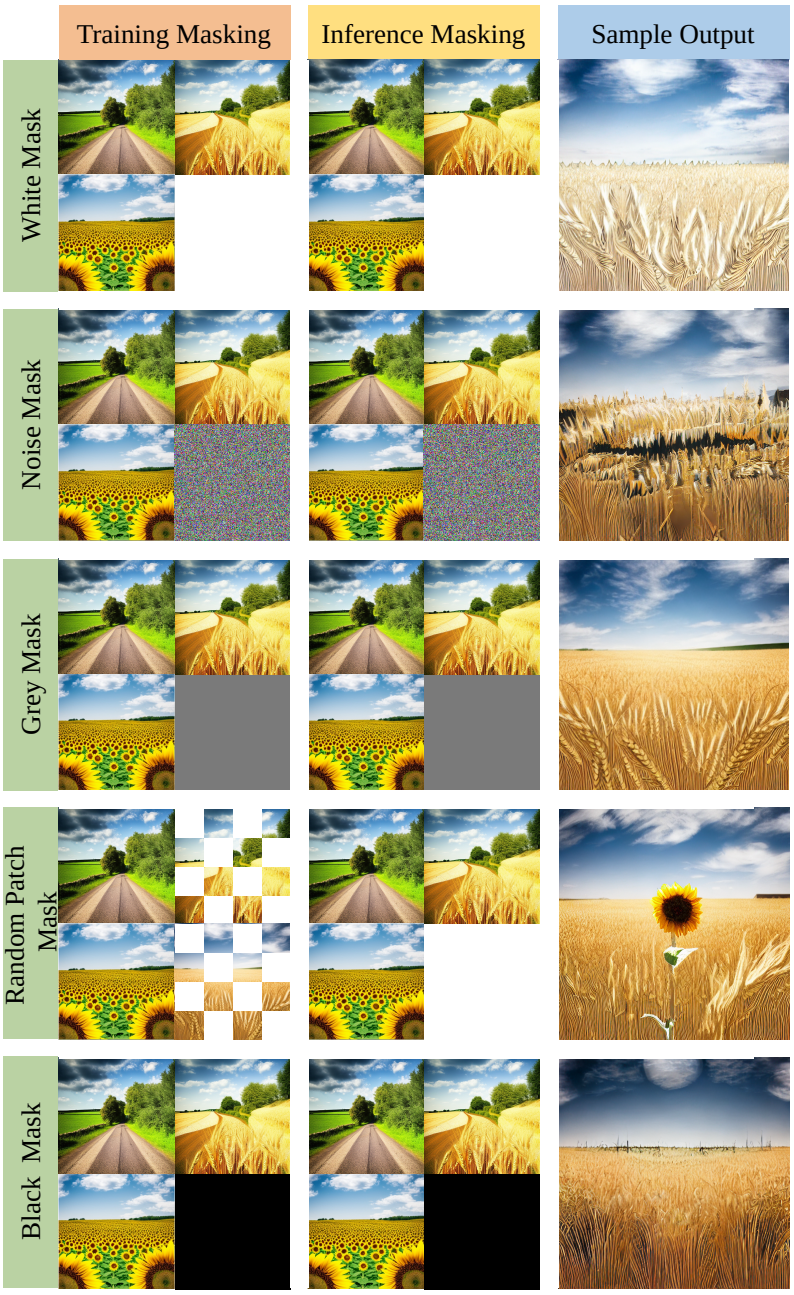

Fig. A1: Performance of InstructGIE with different masking format for visual prompted conditions.

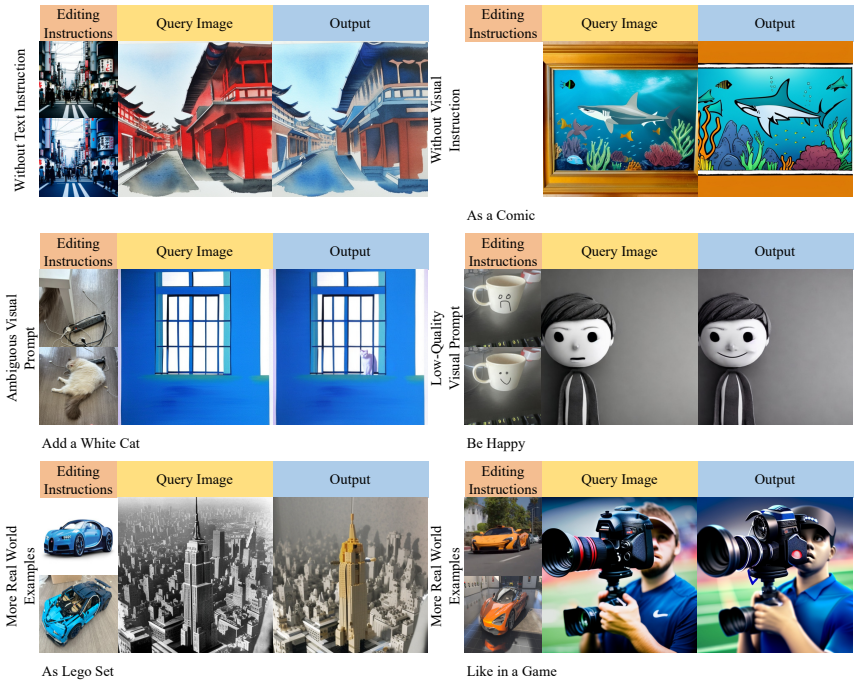

Fig. A2: Additional Edge Case/Real World Visual Performances

C Failure Cases and Potential Future Improvement

We present certain failure cases of our method in Fig. A6. It appears that unsuccessful outcomes often occur when the model does not adequately balance the attention to text and visual editing instructions. The left example of Fig. A6 shows that the model focuses more on text editing instruction (cat) which not only turns Mona Lisa into a cat figure but also turns the pets she holds into a cat. The right example of Fig. A6) shows that the model pays more attention to visual prompts, where we only require the color of clothes to be modified, but the background color is modified as well following visual guidance.

How to balance the attention to text and visual editing instructions is an important key to ensuring stable performance for a generalizable image editing model. In future works, we plan to conduct experiments with several methods including text and visual editing instruction pooling and fusion to deduce a plausible solution to this problem for generalizable image editing.

D Additional Image Editing Result Visualization

In this section, we showcase additional visual manipulation results demonstrating the generalizability and versatility of the InstructGIE framework across a multitude of image editing tasks, as illustrated in Fig. A3, A4, and A5. These qualitative examples show our method's robust performance in a wide range of generalized editing scenarios and directives. This includes the transformation of one object into another, style transfers, and addition or removal of objects. Furthermore, the results show that our method can generate subtle details, such as realistic human face images.

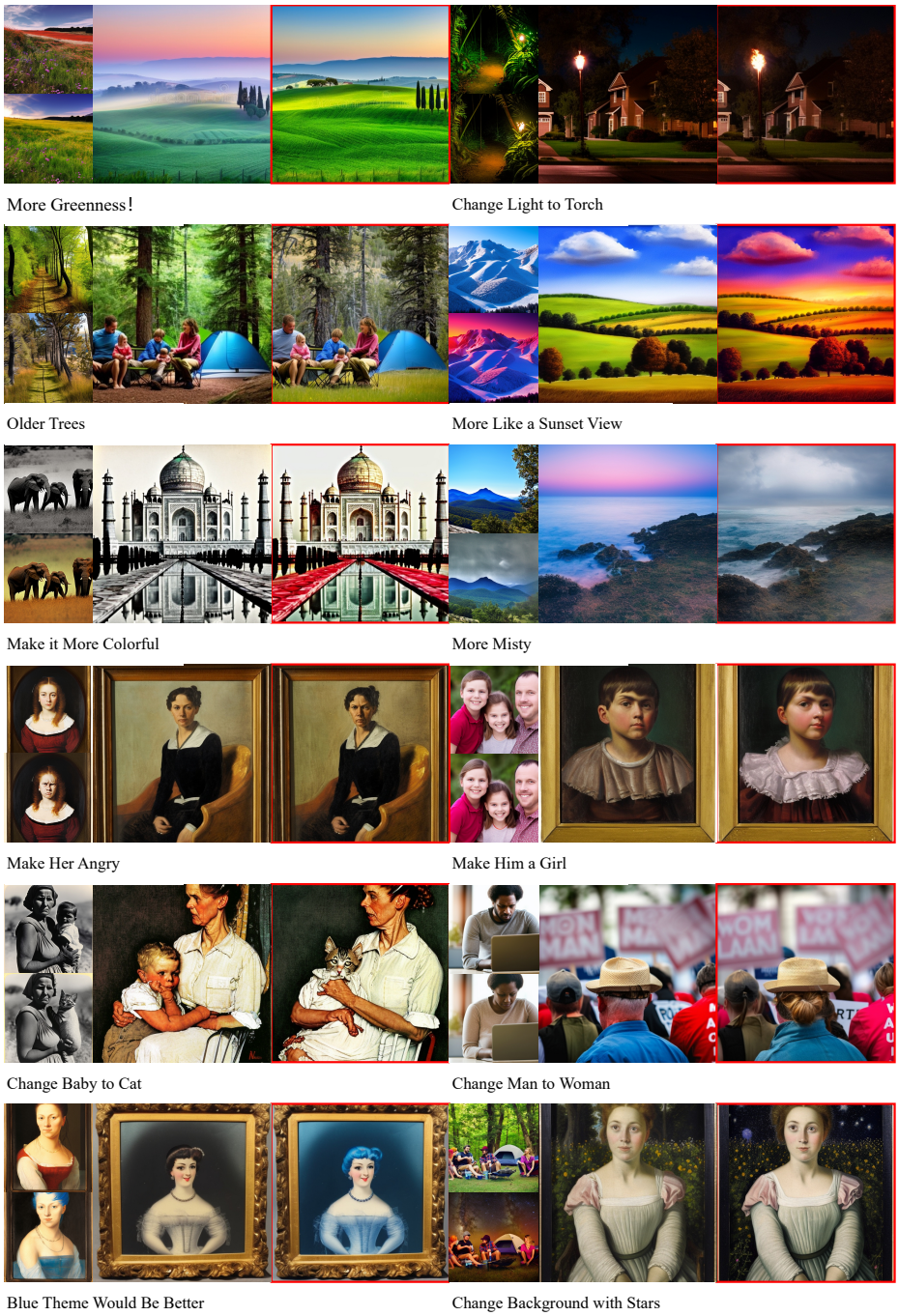

Fig. A3: Additional Image Editing Generalization Results Produced by InstructGIE

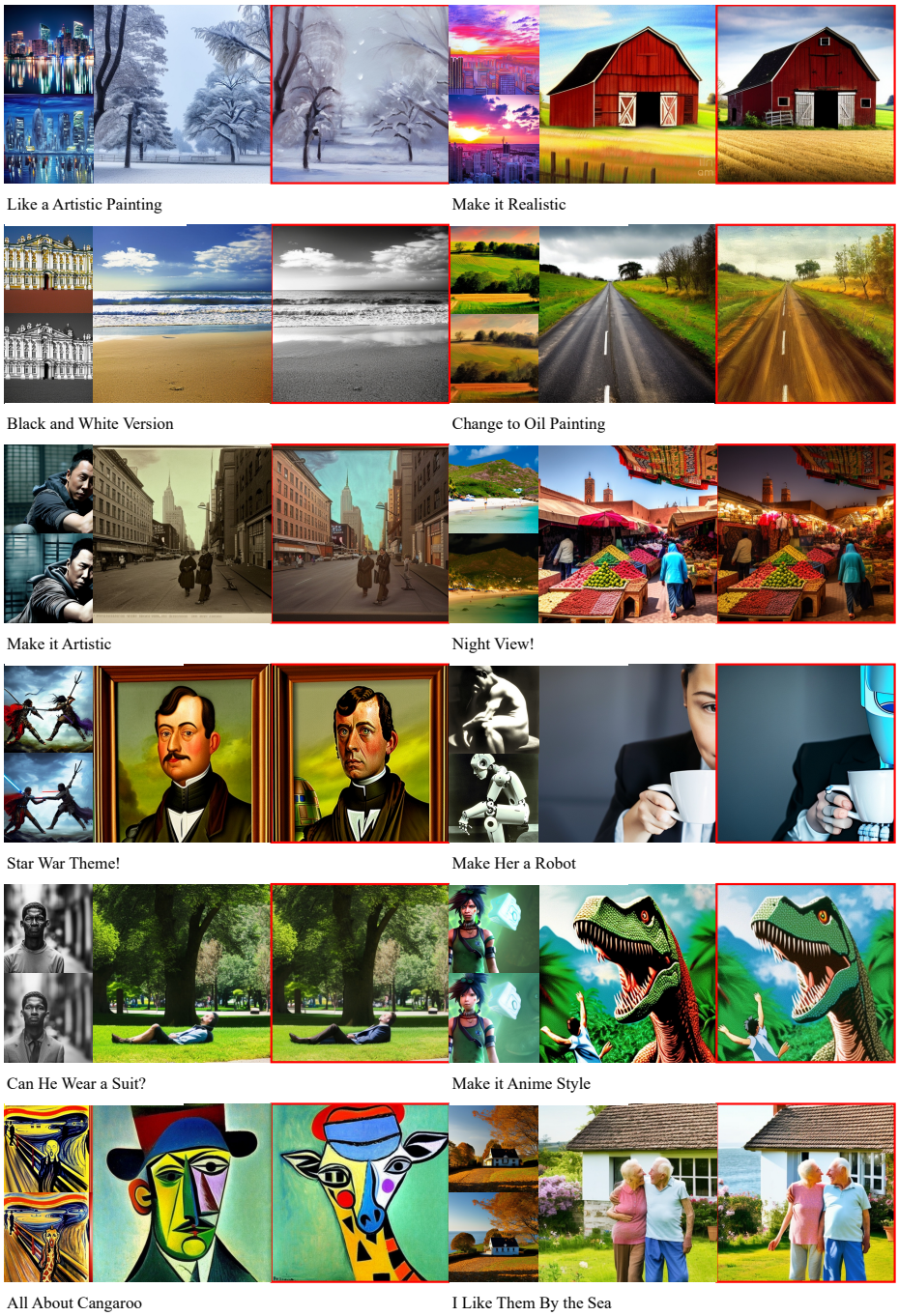

Fig. A4: Additional Image Editing Generalization Results Produced by InstructGIE

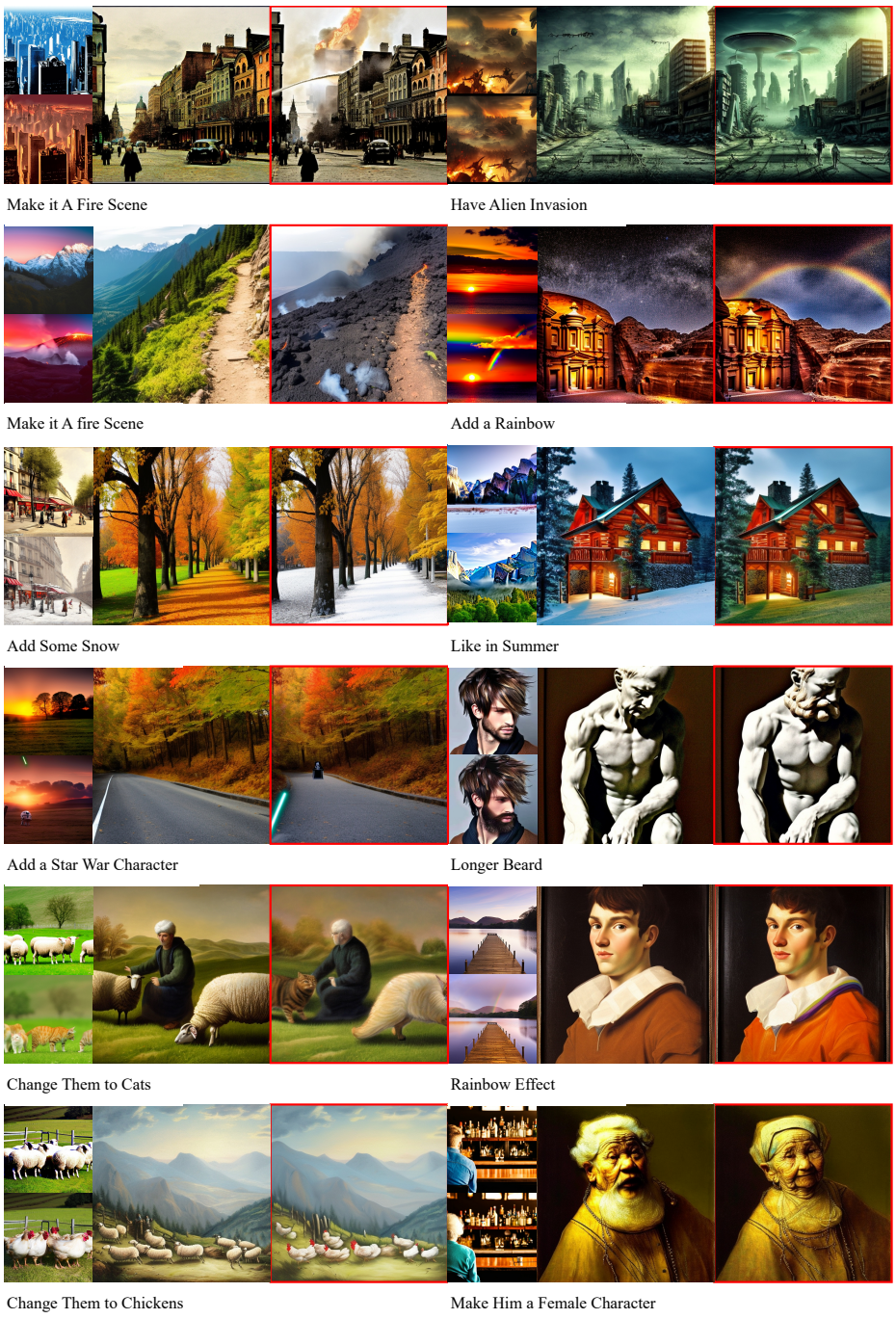

Fig. A5: Additional Image Editing Generalization Results Produced by InstructGIE

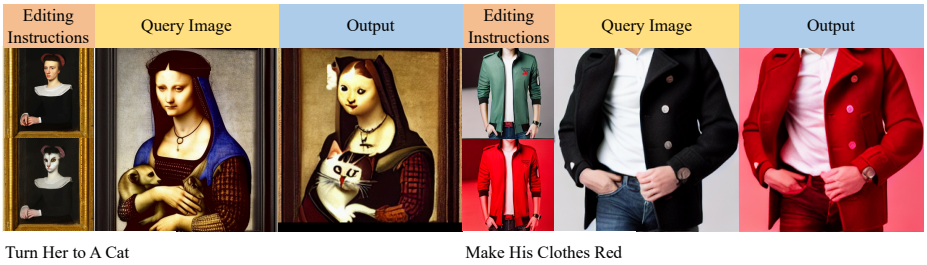

Fig. A6: Selected fail cases from our InstructGIE framework.