A Plug-and-Play Approach for Robust Image Editing in Text-to-Image Diffusion Models

Supplementary Material

A. Comparison of Block-Specific Impact on Diverse Inversion Performance

Qualitative results further confirm that applying RLI to the up block consistently leads to superior editing outcomes compared to the down or mid blocks. Regardless of the editing or inversion method, the up block maintains structural fidelity while achieving the intended semantic changes. This finding is qualitatively supported by the averaged performance across three editing methods (P2P [6], MasaCtrl [2], and PnP [27]), as summarized in Figure 9 for visual comparisons.

B. Analysis of RLI Parameters in Prompt-to-Prompt Editing

To further demonstrate the versatility and effectiveness of our proposed RLI method and to provide detailed insights into its performance under varying configurations, we present additional qualitative results. Figure 10 showcases a range of image reconstruction and prompt-based editing outcomes when applying RLI with Null-text Inversion on SDXL. For each illustrated image, the specific alpha value for RLI, along with the cross-step and self-step values used for the editing process, is explicitly provided. This detailed presentation provides a comprehensive understanding of how different parameter settings affect the final edited image, highlighting RLI's robustness across diverse editing scenarios.

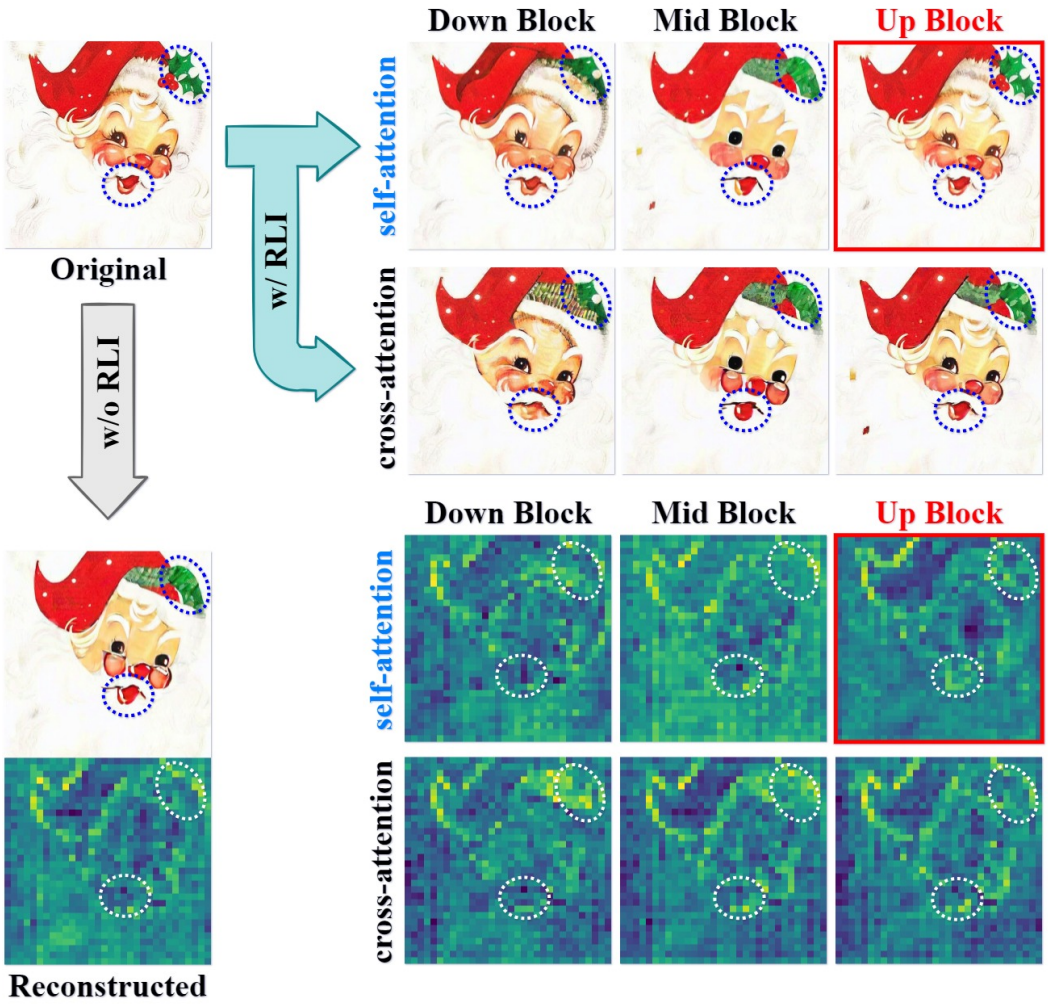

Figure 9. Qualitative analysis of RLI's impact on reconstruction and attention maps across U-Net blocks and attention types. Self-attention in the Up block notably enhances reconstruction quality and attention precision (*e.g.*, chin, hat's flower).

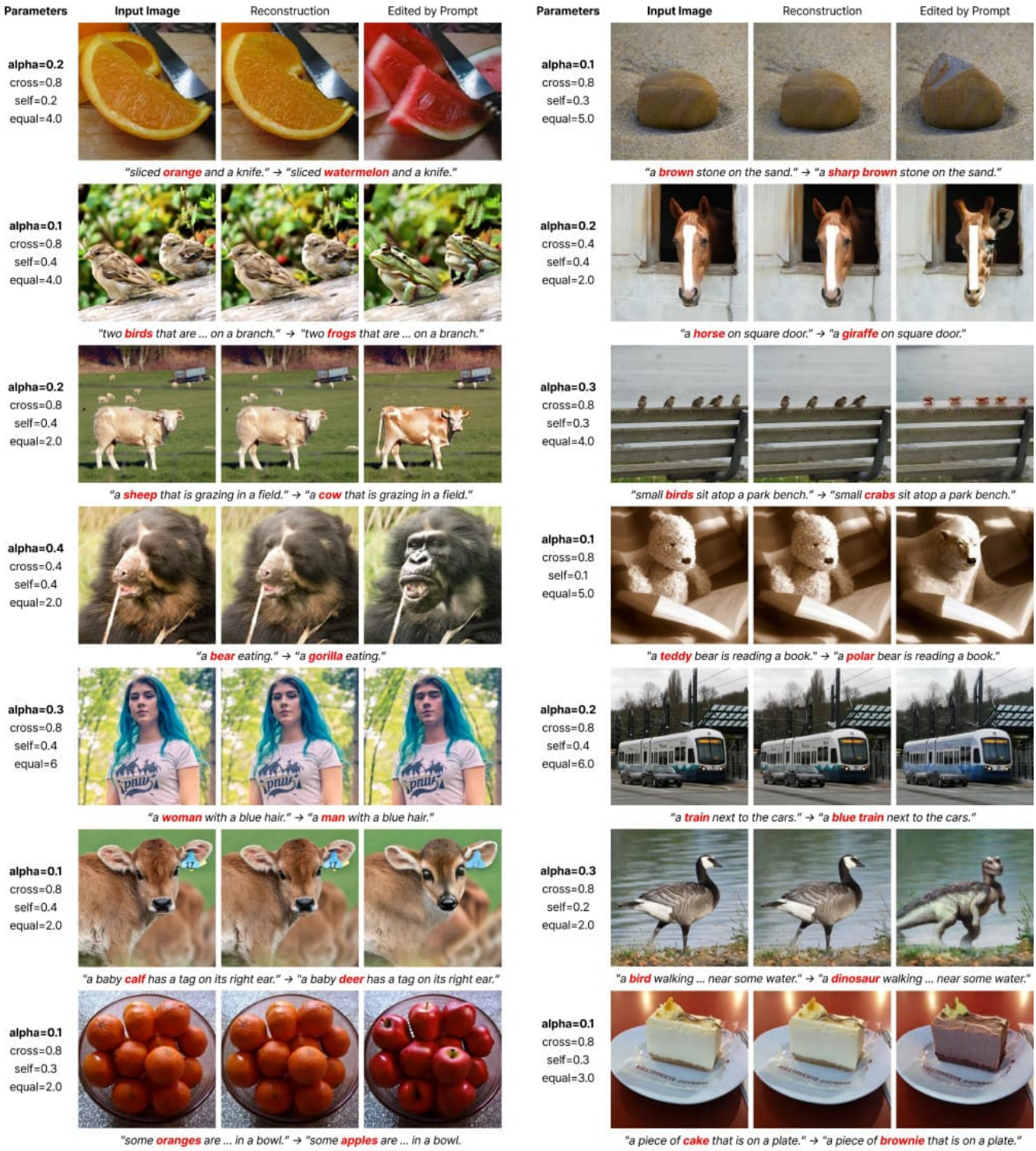

Figure 10. Qualitative results of image reconstruction and prompt-based editing using RLI with Null-text Inversion on SDXL.