

## A. Supplementary Material

### A.1. Evaluation Guide for Human Raters

This section provides the detailed 1–4 criteria used by annotators. All dimensions are strictly independent.

**Instruction Following (IF).** **Core question:** Does the edited video correctly satisfy the removal instruction?

- **4:** Perfect adherence. Target instance is removed cleanly with no obvious remnants.
- **3:** High adherence. Target removed with minor residual imperfections.
- **2:** Low adherence. Target partly removed, or obvious residual traces remain.
- **1:** Failure. Target remains or output is largely unrelated to the instruction.

**Rendering Quality (RQ).** **Core question:** Is the filled content visually plausible and temporally stable?

- **4:** Excellent quality, negligible artifacts, temporally stable.
- **3:** Acceptable quality, moderate artifacts but viewable.
- **2:** Poor quality, obvious distortion/flicker affecting readability.
- **1:** Unusable quality, severe artifacts or physical implausibility.

**Edit Exclusivity (EE).** **Core question:** Are non-target regions preserved?

- **4:** Non-target regions are effectively unchanged.
- **3:** Minor non-target drift but scene remains mostly preserved.
- **2:** Noticeable over-editing in multiple non-target areas.
- **1:** Major uncontrolled edits; scene is substantially altered.