

# DuplexBridge: Coordinating Understanding and Generation for Unified Image Editing

Yidi Liu<sup>1</sup>, Jie Xiao<sup>2</sup>, Chaojie Mao<sup>2</sup>, Jingfeng Zhang<sup>2</sup>, Xueyang Fu<sup>1,\*</sup>

<sup>1</sup>University of Science and Technology of China

<sup>2</sup>Wan Team, Alibaba Group

\*Corresponding author

Complex image editing requires faithful task interpretation and accurate execution, but a failed edit does not reveal which responsibility failed. We present DuplexBridge, which uses a unified model’s own understanding capability to enhance prompts for its own generation capability, without an external understanding module. A source-grounded bridge specifies targets, final-state changes, and preservation constraints. Understanding is supervised on fidelity to the original request without observing the generated image; generation is supervised on execution of the condition it actually receives. Reciprocal trust checks selectively admit global task feedback to each role while retaining independently valid local supervision. We introduce CE-Bench, a 1,200-task benchmark measuring joint satisfaction of task-derived change and preservation requirements. Across three model configurations, the full system improves complete compliance over direct editors by 21.75–25.75 percentage points, with gains also observed on KRIS-Bench, RISEBench, ImgEdit, and GEdit-Bench. From a common supervised Wan-Image initialization, gated joint training reaches 77.08% complete compliance, compared with 70.92% without gates and 75.00% when generation-local supervision instead checks the original task with both gates retained. These results support distinct local responsibilities and selective shared feedback for joint understanding–generation adaptation.

**Date:** October 12, 2026

## 1 Introduction

Complex image editing must bind changes to the correct source content, infer final states, and preserve protected content. Setting only an analog clock to the digital clock’s time plus 90 minutes requires all three.

Image-grounded interpretation and joint optimization already support complex editing (Fu et al., 2024; Huang et al., 2024). We study *native prompt enhancement* (PE) in unified models (Deng et al., 2025; Mao et al., 2026): their own understanding capability, U, constructs conditions for their own generation capability, G, without an external understanding module. When adapting both roles, what should supervise their local responsibilities, and when should they also learn from final task outcomes?

A task-level score alone does not distinguish faithful task expression from faithful execution. In figure 1(a), the digital reading 14:35 gives an analog target of 4:05. U may instead use the wrong reference and specify 11:50, which G faithfully renders. Alternatively, U specifies 4:05 correctly, but G changes the protected digital display. These failures motivate distinct checks of prompt fidelity, received-condition execution, and task outcome.

Executable planning and two-role adaptation have precedents (Zhou et al., 2026; Jiao et al., 2026). DARS routes planner–renderer updates using task-derived checklists and post-render

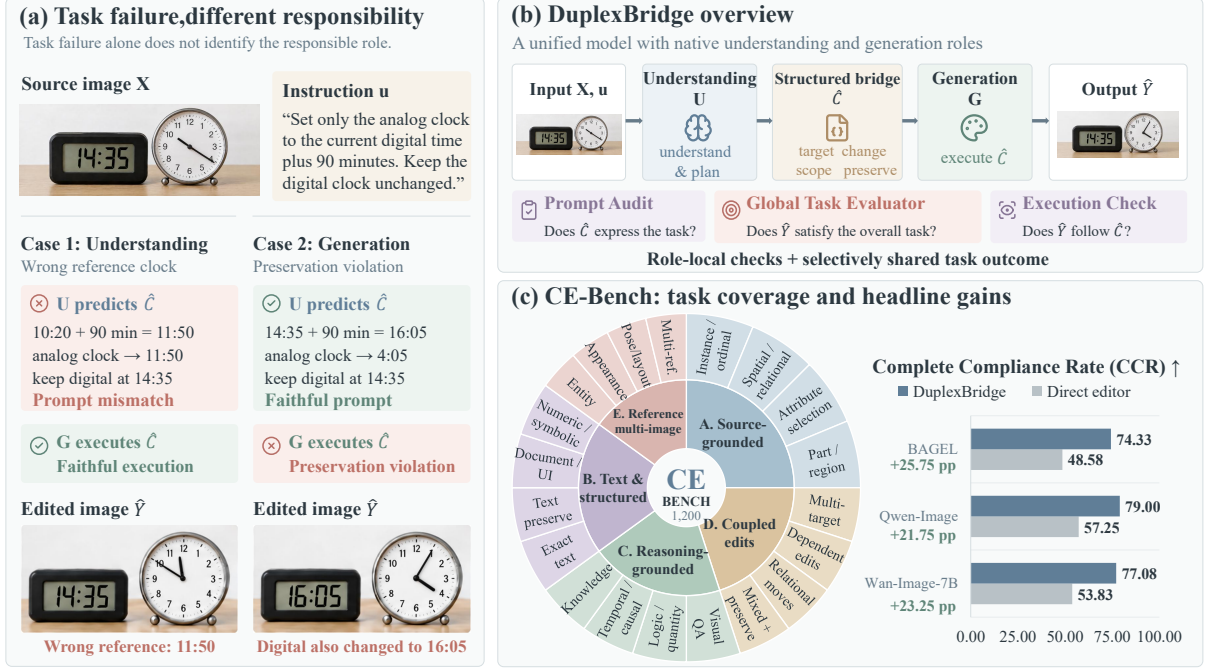


**Figure 1 DuplexBridge overview.** (a) Task failures with different local responsibilities. (b) Native U–G communication and three distinct checks. (c) CE-Bench coverage and full-system CCR gains, including PE, SFT, and RL. Primary sectors encode 300/240/240/240/180 tasks; outer subdivisions illustrate capabilities rather than subtype frequencies.

planner diagnostics (Cao et al., 2026); DDA-Thinker separately evaluates plans and visual outcomes while keeping the editor frozen (Yang et al., 2026). DuplexBridge combines distinct local supervision references for both adapting roles with reciprocal checks that select only their additional global-feedback losses.

The bridge specifies source-bound targets, changes, and preservation (figure 1). G receives only the source and this condition, which also defines G-local execution checks.

U-local checks compare the condition with the task without seeing the generated image. A separate global reward assesses the final edit against the task. Prompt trust gates this reward to G; execution trust gates it to U. Valid local supervision remains when either gate closes. Training alternates role updates; inference follows the native U-to-G path.

CE-Bench requires all task-derived change and preservation checks to pass; these requirements are independent of the predicted bridge. Full-system complete compliance rate (CCR) gains over direct editors span 21.75–25.75 points across three configurations, alongside public-benchmark improvements. From a common supervised Wan-Image initialization, full reaches 77.08% CCR versus 70.92% without gates and 75.00% with original-task G-local and both gates retained. Group-structure-matched random selection reaches 72.25%, supporting semantic selection in this setting.

Our contributions are:

1. A checkable bridge within a unified model that defines distinct local supervision targets for task expression and received-condition execution.
2. A joint learning rule that retains valid local supervision while selectively sharing global task feedback through reciprocal partner-trust gates.
3. CE-Bench, a 1,200-task evaluation of complete editing fulfillment, supported by public benchmarks and controlled comparisons of the feedback design.

## 2 Related Work

**Understanding-generation coordination.** MGIE learns expressive instructions and jointly trains the multimodal language model and image editor (Fu et al., 2024). SmartEdit strengthens complex instruction editing with bidirectional interaction between image features and MLLM outputs (Huang et al., 2024). These works establish understanding-guided editing and coupled adaptation; SmartEdit’s bidirectionality concerns feature interaction, whereas ours concerns training feedback between the understanding and generation roles. BAGEL unifies image understanding and generation; Uni-Edit studies editing-driven adaptation (Deng et al., 2025; Zheng et al., 2026). Unified Thinker uses successive understanding and generation phases, ThinkGen applies SepGRPO, and ThinkRL-Edit decouples the two roles’ updates (Zhou et al., 2026; Jiao et al., 2026; Li et al., 2026). UniGRPO and BRAID further study joint reasoning and synthesis (Liu et al., 2026; Hu et al., 2026). DuplexBridge focuses on the supervision references and feedback eligibility within this co-adaptation: task fidelity, received-condition execution, and task outcome train distinct loss components. Here, *native* understanding means that the unified model uses its own understanding capability for prompt enhancement and passes the result to its own generation capability, without an external understanding module in that inference path. This definition permits external judges and teachers for training supervision.

**Structured prompt enhancement and feedback.** PromptEnhancer optimizes a rewriter with a frozen generator; its published alignment reward evaluates generated images against the rewritten prompt (Wang et al., 2026). Thus, received-condition image assessment is already an established component. Wan-Image and SenseNova-U1.5 use structured editing conditions (Mao et al., 2026; Diao et al., 2026). DDA-Thinker already assesses executable plans without observing generated images and maintains separate cognitive and visual reward streams, while keeping the editor frozen (Yang et al., 2026). DARS addresses the same ambiguity between planning and execution failures. It jointly trains a planner and renderer with an original-task checklist, post-render planner diagnostics, prefix gating across plan fields, and variance-based module weights (Cao et al., 2026). DuplexBridge combines different local references within co-adaptation: U-local judges task fidelity without the rendered image, and G-local judges execution against the condition actually received. Reciprocal partner trust selects only the additional global losses, preserving valid local supervision when global feedback is withheld. This combination makes local responsibility and global-feedback eligibility explicit within the same training procedure. The gates implement selective feedback, not counterfactual credit estimation (Foerster et al., 2018).

**Reward modeling and editing evaluation.** SLD detects and corrects generation errors through an inference-time loop (Wu et al., 2024); DuplexBridge uses execution checking to determine which feedback can train each role. RewardDance uses comparative token probabilities; UnifiedReward-Think and Z-Reward study reasoning-enhanced and distilled judgment (Wu et al., 2025a; Wang et al., 2025; Jin et al., 2026). Our distilled RM supports the shared outcome signal through direct pairwise scoring against fixed reference edits. TIFA and Davidsonian Scene Graph motivate decomposed visual checking (Hu et al., 2023; Cho et al., 2024). ImgEdit, GEdit, RISEBench, and KRIS-Bench assess complementary dimensions of editing and reasoning (Ye et al., 2025; Liu et al., 2025b; Zhao et al., 2025; Wu et al., 2025b). CE-Bench emphasizes *complete* compliance with source-grounded changes and preservation constraints. Its requirements come from the original task, independently of a model’s predicted bridge, and evaluation accepts final images.

## 3 A Structured Bridge for Understanding and Generation

### 3.1 Native roles and the forward path

Let  $X$  contain the source image and any task-provided reference images, with their roles identified, and let  $u$  be the editing instruction. DuplexBridge uses a unified model’s own understanding role

$U_\theta$  for prompt enhancement (PE) and its own generation role  $G_\phi$  for editing:

$$(z_\theta, \hat{C}) \sim \pi_\theta(\cdot \mid X, u; \psi), \quad \hat{Y} \sim G_\phi(\cdot \mid X, \hat{C}; \psi). \quad (1)$$

Here  $z_\theta$  is internal interpretation in a `<think>` block and  $\hat{C}$  is the public structured bridge. Trainable role parameters  $\theta, \phi$  are disjoint; fixed pretrained parameters  $\psi$  may include a shared backbone. No external understanding module constructs  $\hat{C}$ . Generation receives  $X$  and  $\hat{C}$  in a fresh context, without the original instruction or U’s internal reasoning.

### 3.2 Making the editing requirements explicit

U resolves target identities, reasoning-dependent answers, and terminal states before constructing  $\hat{C}$ . The bridge binds changes to source entities, separates editable from protected content, and specifies preservation requirements and forbidden effects. Entailed requirements are made explicit without additional edits; requested text retains its exact characters. CE-v1 organizes these requirements as grouped JSON fields `goal`, `scope`, `ops`, `preserve`, and `avoid` (figure 2). The goal summarizes the operative fields without adding obligations.

The public condition also defines the execution checks:

$$D(\hat{C}) = (D_\Delta, D_{\text{inv}}), \quad Q_{\text{exec}} = \mathcal{A}(\hat{C}), \quad (2)$$

where  $D_\Delta$  and  $D_{\text{inv}}$  are disjoint change and invariant groups. A source resolver binds targets in  $X$ ; fixed templates check applicable initial states before output requirements. Repeated obligations are deduplicated; overlapping preservation scopes are partitioned or deduplicated after semantic validation. Template enumeration is deterministic; source binding and visual judgment remain semantic operations.

The compiler uses the same condition as G and cannot recover omitted requirements from  $u$  or  $z_\theta$ . A clearly specified edit can therefore be executable yet wrong for the user’s request, motivating separate prompt and execution feedback. U is initialized on accepted teacher responses containing internal interpretation and public JSON before RL. Sections 3.3 and 6.2 describe the bridge semantics and the initialization controls.

### 3.3 Source binding, coverage, and validity

The bridge supports recoloring, replacement, insertion, removal, movement, text transformation, and observable state changes. It expresses the requested terminal state, including exact final text or a reasoning-derived value, rather than an unresolved instruction to reason. Editable and protected source content have disjoint roles. Targets must resolve to a source instance or a determinate set; matching the edited output retains that source identity. Sharing a description does not guarantee that U, G, and a visual judge interpret it identically.

Each execution group conjoins source grounding, applicable initial-state checks, and the corresponding output requirement. A failed or unresolved prerequisite marks its group as non-passing once; training eligibility is assessed separately. Preservation requirements reuse source evidence and are deduplicated or partitioned when scopes overlap, so repeated wording does not increase reward weight. These reduction rules also apply to public evaluation, whose requirements come from the original task rather than the predicted bridge.

In the clock example, the source analog display reads 10:20 and the protected digital display reads 14:35. The requested digital time plus 90 minutes yields 16:05, or 4:05 on the analog dial. The bridge specifies 4:05 as the analog terminal state; reading the digital display supplies evidence without making it an edit target. Its unchanged time and appearance, the analog dial and case, and the surrounding scene remain preservation obligations.

The goal summary must be entailed by the operative requirements. Unsupported operations, unresolved targets, contradictory clauses, or a goal containing an otherwise omitted obligation

cannot silently produce a valid execution specification. Invalid or unresolvable specifications and uncertain execution judgments are excluded from execution training, even when diagnostics record a failed group. Confident failures on valid, resolvable checks remain learning signals. A confident prompt-fidelity judgment may still supervise U locally. Private interpretation, expected judge answers, benchmark identifiers, and rewards are excluded from the public condition passed to G.

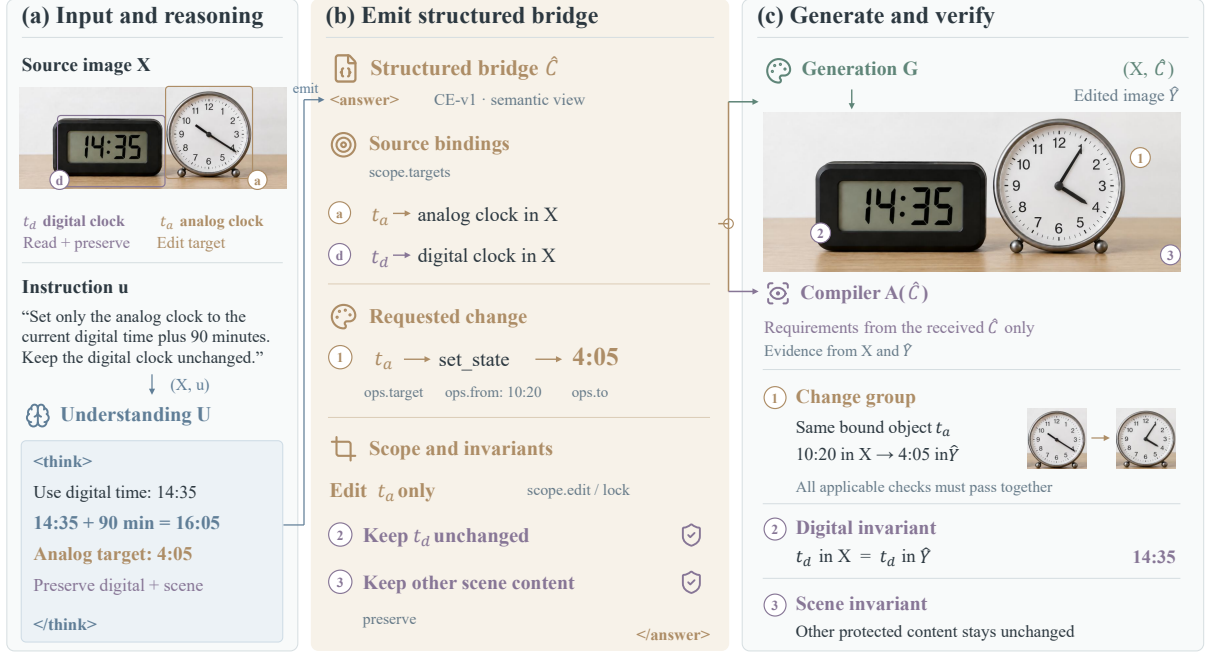


**Figure 2 From understanding to a checkable bridge.** U resolves the clock task and separates the editable analog clock from protected digital and scene content. Only the public condition enters G and the execution compiler; change and invariant checks use that same condition.

## 4 Coordinating Learning with Local and Global Feedback

DuplexBridge separates three evaluations by the information they receive: prompt fidelity uses  $(X, u, \hat{C})$ , execution uses  $(X, \hat{C}, \hat{Y})$ , and the global outcome uses  $(X, u, \hat{Y})$  (figure 3). Local rewards train each role’s responsibility; partner trust determines when the joint outcome also supervises that role.

### 4.1 Role-local rewards

$$R_U^{\text{loc}} = F_U(X, u, \hat{C}), \quad R_G^{\text{loc}} = F_G(X, \hat{C}, \hat{Y}). \quad (3)$$

**Understanding: prompt fidelity.** A separately validated reference  $C^*$  supports format, requirement-level semantic precision/recall, and consistency/grounding judgments:

$$R_U^{\text{loc}} = s_C = (R_{\text{format}} + R_{\text{plan}} + R_{\text{clear}})/3. \quad (4)$$

The plan score uses maximum one-to-one semantic matching between predicted and verified requirements: target identity, operation, terminal value, and invariant scope must agree. With  $T$  matches among  $P$  predicted and  $R$  reference groups,  $R_{\text{plan}} = 2T/(P + R)$ . Format is a binary schema-conformance check; clarity averages a no-contradiction judgment and the resolved-source-binding fraction, with malformed or empty predictions scored zero. This check never reads  $\hat{Y}$ . Unresolved judgments abstain. The verified reference is distinct from teacher SFT targets (section 3.2).

**Generation: execution and local quality.** For  $Q_{\text{exec}} = \{Q_1, \dots, Q_K\}$ , each change group combines target grounding with its visual postconditions, while each distinct invariant forms another group. With visual judge  $V$  and check thresholds  $\tau_q$ ,

$$R_{\text{exec}} = \frac{1}{K} \sum_{k=1}^K \prod_{q \in Q_k} \mathbf{1}[V(X, \hat{Y}, q) \geq \tau_q], \quad (5)$$

$$R_G^{\text{loc}} = w_{\Delta} R_{\text{exec}, \Delta} + w_{\text{inv}} R_{\text{exec}, \text{inv}} + w_q R_{\text{quality}}. \quad (6)$$

The nonnegative weights sum to one; change and invariant scores average their respective groups, both of which must be nonempty for execution training. Quality excludes instruction compliance, and preservation is counted once. Invalid or unresolvable specifications and uncertain execution judgments are excluded by the eligibility mask; low scores on valid, confidently judged checks remain learning signals (section 3.3).

## 4.2 Global task feedback

A frozen pairwise reward model (RM) compares  $\hat{Y}$  with baseline edits from a fixed, nonempty set  $\mathcal{R}_{X,u}$  under the original instruction and rubric  $i$ . It receives neither  $\hat{C}$  nor U’s reasoning. Following the token-probability interface of RewardDance (Wu et al., 2025a),

$$R_{\text{global}} = F_{\text{global}}(X, u, \hat{Y}; \mathcal{R}_{X,u}) = \frac{1}{|\mathcal{R}_{X,u}|} \sum_{Y_r \in \mathcal{R}_{X,u}} P_{\omega}(\text{Yes} \mid X, u, \hat{Y}, Y_r, i). \quad (7)$$

Here **Yes** favors  $\hat{Y}$ ; its raw vocabulary probability gives a relative preference score in  $[0, 1]$ , not an absolute completion probability. References are independent of candidate bridges and shared across each rollout group. The direct-scoring RM distills a thinking teacher’s judgment distribution and remains frozen during U/G training; its formulation and evaluation are in sections 4.5 and 6.8.

## 4.3 Bidirectional gates

Prompt trust controls G-global feedback; execution trust controls U-global feedback. Let  $c_{\text{audit}}$  indicate an available verified  $C^*$  and a sufficiently reliable prompt-fidelity judgment, and let  $b_E, b_Q, b_F$  indicate usable execution, quality, and global judgments. For a valid specification, prompt trust requires

$$v_C = \mathbf{1}[R_{\text{format}} = 1] \mathbf{1}[R_{\text{plan}} \geq \tau_p] \mathbf{1}[R_{\text{clear}} \geq \tau_c] \times \mathbf{1}[\text{required targets resolved}]. \quad (8)$$

Here  $\tau_p, \tau_c$  are the minimum plan and clarity scores; invalid specifications have  $b_E = v_C = 0$ . Unresolved source bindings or uncertain execution judgments set  $b_E = 0$ ; unresolved required targets also give  $v_C = 0$ . Judgment usability is separate from success:  $c_{\text{audit}}, b_E, b_Q$  require reliable judgments, with reliability estimators and thresholds fixed using independent development labels. The reciprocal gates and global-loss eligibility are

$$\begin{aligned} m_{U \rightarrow G} &= c_{\text{audit}} v_C, & e_G^{\text{glob}} &= b_F m_{U \rightarrow G}, \\ m_{G \rightarrow U} &= b_E \mathbf{1}[R_{\text{exec}} \geq \tau_e], & e_U^{\text{glob}} &= b_F m_{G \rightarrow U}. \end{aligned} \quad (9)$$

Local eligibility is independent of these gates:  $e_U^{\text{loc}} = c_{\text{audit}}$  and  $e_G^{\text{loc}} = b_E b_Q$ . Thus  $b_E = 0$  excludes G-local and U-global, while a reliable prompt judgment can still train U locally; diagnostic zeros for unusable checks do not enter the loss. Judgment usability differs from task success: a confidently identified failure can supply a low local reward. Gate-only ablations bypass semantic trust while retaining judgment-confidence and availability tests, so an unavailable check is never treated as a successful execution.

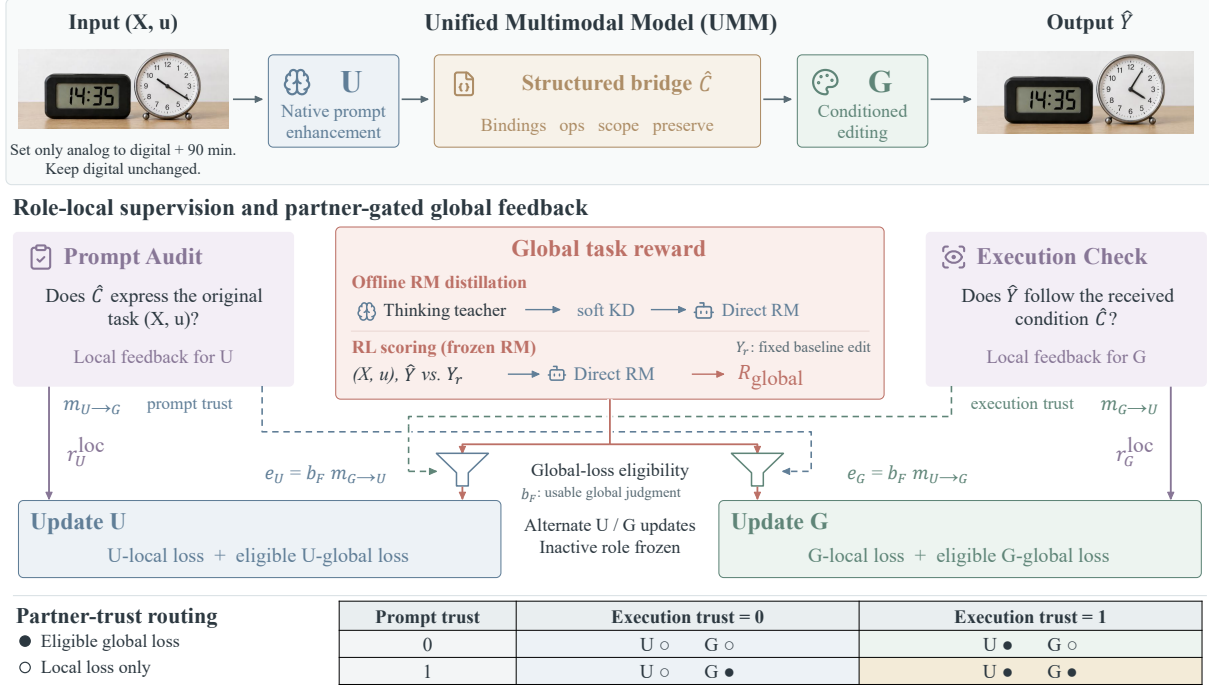


**Figure 3 Bidirectional partner-gated feedback.** Execution trust gates U-global; prompt trust gates G-global. Local feedback retains independent validity masks. The diagram abbreviates  $e_U^{\text{glob}}, e_G^{\text{glob}}$  as  $e_U, e_G$  and local rewards as  $r_U^{\text{loc}}, r_G^{\text{loc}}$ . The routing matrix assumes usable judgments.  $Y_r$  is a baseline edit, distinct from task-provided reference images.

In figure 1(a), if G faithfully sets the analog clock to U’s incorrect 11:50, G retains local supervision. Passed execution trust admits global feedback to U; failed prompt trust excludes G-global. When U correctly specifies 4:05 but G changes the protected digital clock, passed prompt trust admits G-global; U-global is excluded if execution trust fails. The mean threshold can still accept a missed change or preservation failure (section 4.4).

#### 4.4 Alternating updates with four loss components

Each role combines two separately normalized group-relative clipped losses:

$$\begin{aligned}\mathcal{L}_U &= \mathcal{L}_{\text{GR}}^{U, \text{loc}}(R_U^{\text{loc}}; e_U^{\text{loc}}) + \lambda_U \mathcal{L}_{\text{GR}}^{U, \text{glob}}(R_{\text{global}}; e_U^{\text{glob}}), \\ \mathcal{L}_G &= \mathcal{L}_{\text{GR}}^{G, \text{loc}}(R_G^{\text{loc}}; e_G^{\text{loc}}) + \lambda_G \mathcal{L}_{\text{GR}}^{G, \text{glob}}(R_{\text{global}}; e_G^{\text{glob}}).\end{aligned}\quad (10)$$

For U updates, freeze G, sample complete U responses, and generate several images per response. Average global scores over eligible images within each response, then weight eligible responses equally; a response with none can still contribute to U-local. For G updates, freeze U and compare stochastic image trajectories conditioned on one bridge.

For component  $t \in \{U_{\text{loc}}, U_{\text{glob}}, G_{\text{loc}}, G_{\text{glob}}\}$ , let  $I_t = \{i : e_{ti} = 1\}$  and compute its reward mean  $\mu_t$  and standard deviation  $\sigma_t$  only on  $I_t$ . For an eligible, nondegenerate group,

$$\begin{aligned}A_{ti} &= \frac{R_{ti} - \mu_t}{\sigma_t + \epsilon}, \\ \mathcal{L}_t &= \frac{1}{|I_t|} \sum_{i \in I_t} [-\min\{\rho_{ti} A_{ti}, \text{clip}(\rho_{ti}, 1 - \epsilon, 1 + \epsilon) A_{ti}\} + \beta_t \mathcal{D}_{ti}].\end{aligned}\quad (11)$$

Here  $\rho_{ti}$  is the current-to-behavior policy likelihood ratio and  $\mathcal{D}_{ti}$  the frozen-reference policy penalty. Each component normalizes rewards and applies its KL penalty only on its eligible set; replacing excluded rewards by zero inside shared normalization is not equivalent. Groups with fewer than two eligible samples or zero reward variance are skipped for that component, including

KL. U uses text-policy likelihoods and G uses stochastic flow trajectory likelihoods (Shao et al., 2024; Liu et al., 2025a). A sample excluded from global feedback can retain local supervision; positive raw reward need not imply positive centered advantage.

**Execution trust and critical requirements.** The execution gate tests the average group score. With one failure among  $K$  groups,  $R_{\text{exec}} = (K - 1)/K$  can exceed  $\tau_e$  even when a requested change or preservation obligation fails. Deduplication prevents repeated obligations from increasing their weight, but distinct invariants can still dilute one failure. Execution trust therefore does not require every critical obligation to pass.

**Conditional return and selection.** U-global averages accepted images within each candidate and weights eligible candidates equally. It therefore estimates return conditional on acceptance. With  $L$  conditionally i.i.d. image-judgment samples and per-image acceptance probability  $p(C)$ , a candidate is included with probability  $1 - (1 - p(C))^L$ , before group-size and variance filtering. Equal weighting does not undo this selection, and accepted-return rankings can differ from rankings that also account for rejected images. The gates define selective objectives, not unbiased unconditional-return gradients or counterfactual causal attribution.

#### 4.5 Reasoning-distilled global judgment

For pairwise input  $x = (X, u, Y_A, Y_B, i)$ , a frozen thinking teacher produces reasoning  $z$  and a judgment distribution  $\mathbf{q}_T(x, z)$ . The direct student predicts  $\mathbf{p}_\omega(x)$  from the same visual inputs, candidate order, and rubric, without receiving or decoding the teacher’s reasoning. Both compare original-task fulfillment, preservation, and visual quality. Predicted bridges, private U reasoning, human preference labels, and test annotations are excluded from their inputs.

The RL score uses raw vocabulary probability  $P_{\text{Yes}}$ , whereas the binary evaluation probability conditions on the two judgment tokens:

$$r_{\text{raw}} = P_{\text{Yes}}, \quad r_{\text{bin}} = \frac{P_{\text{Yes}}}{P_{\text{Yes}} + P_{\text{No}}}. \quad (12)$$

For raw-score distillation, the common outcome space retains the masses for Yes, No, and all other vocabulary tokens. Matching only binary-normalized judgments would define a different target. Soft judgment distillation optimizes

$$\mathcal{L}_{\text{KD}} = \mathbb{E}_{x,z} D_{\text{KL}}(\text{sg}[\mathbf{q}_T(x, z)] \| \mathbf{p}_\omega(x)), \quad (13)$$

with only the student updated. There is no loss on teacher reasoning tokens. In expectation, minimizing this loss targets the frozen teacher’s marginal judgment distribution  $\mathbb{E}_z[\mathbf{q}_T(x, z)]$ . It transfers judgments rather than the teacher’s internal reasoning algorithm and does not guarantee calibration. After distillation, the RM stays frozen during U/G training and supplies only global feedback. Independent preference labels, candidate-order swaps, and difficult-pair evaluation assess its reliability (section 6.8).

## 5 Experimental Setup

### 5.1 Models and comparison scope

Image editing uses BAGEL, Qwen-Image-Edit-2511 (Qwen-2511), and Wan-Image-7B, with mechanism ablations on Wan-Image. Each model supplies its own U and G for native PE. Qwen-Image labels in the figures denote Qwen-2511. Direct baselines edit the original instruction; full systems add native PE, structured supervised initialization, and joint RL.  $B_0$  denotes the common supervised initialization before RL. U-only, G-only, local-only, and joint variants are independently trained from  $B_0$ , with common supervision and reference access. Unless varied

explicitly, global feedback uses the reasoning-distilled RM. RM replacement experiments use equal update steps and rollout budgets. The controls distinguish complete-system gains from supervised initialization, communication, and RL effects (sections 6.2 to 6.4).

## 5.2 CE-Bench and complete task fulfillment

Our 1,200-task benchmark tests whether target binding, terminal-state changes, and preservation constraints are jointly satisfied. Section 5.3 gives its five-category taxonomy. Evaluation uses the original task, source, required reference images, and final output; requirements are independent of the model’s predicted bridge. For item  $n$ , let  $a_{nk}$  denote a binary pass for requirement group  $k$ , and let  $P_n$  be the mean pass rate over its preservation groups. We report

$$AC = \frac{1}{N} \sum_n \frac{1}{K_n} \sum_k a_{nk}, \quad CCR = \frac{1}{N} \sum_n \prod_k a_{nk}, \quad Pres = \frac{1}{N} \sum_n P_n. \quad (14)$$

All metrics are percentages and weight items equally; failed outputs remain in the denominator. CCR requires complete compliance with every change and preservation group. A change group includes target grounding and the required terminal state; each distinct preservation obligation counts once. Correctly changing the analog clock but altering the protected digital display fails CCR despite passing the change check. These task-derived groups differ from the predicted-condition groups used for G-local training.

Public obligations are derived from the source, task-provided references, and original instruction before predictions are assessed. Independently verified  $C^*$  supports prompt checks; accepted teacher responses instead supply SFT targets and are not automatically evaluation oracles.

## 5.3 Taxonomy and task allocation

CE-Bench assigns each of its 1,200 tasks one primary category and a secondary capability category (table 1). CE-v1 in section 3.3 names the prompt schema, not a benchmark version.

**Table 1 CE-Bench taxonomy.** The five primary categories are mutually exclusive; the secondary categories describe capability coverage.

| Primary category                     | $N$ | Secondary categories                                                                                                                                                    |
|--------------------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. Source-Grounded                   | 300 | A1 Instance / Ordinal Grounding; A2 Spatial / Relational Grounding; A3 Attribute-Conditioned Selection; A4 Part / Region Grounding.                                     |
| B. Text & Structured Content         | 240 | B1 Exact Text Editing; B2 Multi-Text & Text Preservation; B3 Document / UI Layout Editing; B4 Numeric / Symbolic / Structured Visual Editing.                           |
| C. Reasoning-Grounded                | 240 | C1 Visual Question-to-Edit; C2 Logical / Spatial / Quantitative Reasoning; C3 Temporal / Causal / Physical Inference; C4 Knowledge / Conceptual / Procedural Reasoning. |
| D. Coupled & Compositional           | 240 | D1 Multi-Target Composition; D2 Dependent / Coupled Edits; D3 Relational Rearrangement; D4 Mixed Operations + Preservation.                                             |
| E. Reference-Conditioned Multi-Image | 180 | E1 Entity / Subject Transfer; E2 Attribute / Appearance Transfer; E3 Pose / Layout / Structural Transfer; E4 Multi-Reference Composition & Reasoning.                   |

**Primary-label rules.** Precedence is  $E > D > C > B > A$ . E requires an additional task-provided reference image; otherwise D requires multiple principal changes or coupled editing obligations, C requires inference to derive the target or terminal state, B centers on exact text or structured content, and A on directly visible source grounding. Thus, a supplied replacement number is B, while a number that must be calculated is C; directly stated spatial relations are A, while relations requiring inference are C. Ordinary preservation alone does not make a single-change task D. Nonexclusive tags record grounding, reasoning, operations, reference use, and preservation without counting tasks again in aggregate scores.

**Source and reference roles.** Each input  $(X_{\text{src}}, \mathcal{I}_{\text{ref}}, u)$  requires one edited source image. A–D have no additional reference images; E uses references for entity, appearance, or structural information, with E4 requiring at least two. The source defines the editable scene and protected content, while stable image identifiers bind referenced attributes to their input images. Editors and evaluators receive the same task images in the same order with these roles preserved. These conditioning images are distinct from the baseline reference *edits* used as comparison outputs by the pairwise RM.

AC, CCR, and Pres. follow [equation \(14\)](#). Overall scores weight category means by their item counts (25%, 20%, 20%, 20%, 15% for A–E), with failed outputs retained in their assigned category and denominator.

## 5.4 Public benchmark protocols

We report the KRIS-Bench composite, RISEBench-360 completion accuracy, ImgEdit’s 1–5 score, and GEdit-Bench English Full-set  $G_O$  (0–10) ([Wu et al., 2025b](#); [Zhao et al., 2025](#); [Ye et al., 2025](#); [Liu et al., 2025b](#); [Cao et al., 2026](#)). Cross-system comparisons evaluate complete configurations; Wan-Image ablations test components within one configuration.

RISEBench-360 has 85 temporal, 90 causal, 100 spatial, and 85 logical tasks. Overall accuracy weights these task counts; it is not the unweighted mean of the four category percentages or the continuous visual-dimension scores.

KRIS measures knowledge-dependent editing through a composite, not complete success accuracy. GEdit EN follows the English Full-set  $G_O$  column in DARS Table S5, rather than the intersection of valid outputs across methods. GEditBench v2 uses a different pairwise/Elo protocol ([Jiang et al., 2026](#)); its scores and subsets are not interchangeable with v1. Metrics are reported on their native scales without a cross-benchmark average.

# 6 Experiments

## 6.1 End-to-end editing performance

DuplexBridge improves CE-Bench CCR by 25.75, 21.75, and 23.25 points on BAGEL, Qwen-2511, and Wan-Image, respectively, alongside preservation ([table 2](#)). These full-system gains include PE, supervised initialization, and RL.

On public benchmarks ([table 3](#)), Wan-Image reaches 80.13/29.17 KRIS/RISE, gaining 19.29/16.11 points over its direct baseline. Qwen-2511 reaches 82.16/31.11, versus DARS’s 80.72/27.50. These are complete-system comparisons; component controls follow. Task-specific text-to-image retraining is evaluated on GenEval2, R2I, WISE-V, and DPG in [section 6.10](#).

**Qualitative behavior.** The CE-Bench examples in [figure 4](#) illustrate image-grounded answers, spatial placement, and preservation. DuplexBridge inserts “Jackrabbit” and “45” while retaining the worksheet diagrams, places a sliced piece on the upper-left plate, and combines a man, puppy, and red-patterned toy rabbit with the woman’s base scene. These examples accompany the quantitative results in [table 2](#). [Section 7](#) adds reasoning and physical transformations on RISEBench and KRIS-Bench, as well as additional CE-Bench edits.

## 6.2 Supervised initialization controls

The main tables compare unadapted baselines with full DuplexBridge; the supervised  $B_0$  checkpoint uses native PE and structured SFT, before RL. [Table 5](#) reports all three supervised checkpoints on CE-Bench and public benchmarks. From  $B_0$  to full DuplexBridge, CE CCR increases by 17.91, 16.92, and 17.66 percentage points for BAGEL, Qwen-2511, and Wan-Image-7B. All three configurations improve the four public metrics and each RISE category over their

**Table 2 CE-Bench comparison.** A–E are category CCR (table 1); scores are percentages. Full systems include native PE, structured SFT, and gated RL. Dark/light pink marks the best/second-best values within each model group, including ties.

| Model                          | A            | B            | C            | D            | E            | AC ↑         | Pres. ↑      | CCR ↑        |
|--------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| <b>Commercial Models</b>       |              |              |              |              |              |              |              |              |
| GPT-Image-1.5                  | <b>83.00</b> | <b>78.33</b> | 72.92        | 82.50        | 68.89        | 93.47        | 94.12        | 77.83        |
| Gemini-3-pro-image-preview     | 82.33        | 77.08        | <b>81.67</b> | <b>84.17</b> | <b>72.78</b> | <b>94.08</b> | 93.87        | <b>80.08</b> |
| Seedream 4.5                   | 80.33        | <b>77.50</b> | 62.92        | 72.50        | 58.89        | 91.36        | <b>94.43</b> | 71.50        |
| Gemini-2.5-Flash-Image         | 73.00        | 69.58        | 64.17        | 70.83        | 59.44        | 89.67        | 90.72        | 68.08        |
| <b>Generation-only Editors</b> |              |              |              |              |              |              |              |              |
| FLUX.1-Kontext-dev (direct)    | 58.33        | 43.33        | 30.42        | 43.75        | 33.33        | 75.62        | 86.31        | 43.08        |
| FLUX.2-klein-9B (direct)       | <b>75.67</b> | 64.58        | 47.08        | 60.00        | <b>50.56</b> | <b>85.92</b> | 92.87        | <b>60.83</b> |
| LongCat-Image-Edit (direct)    | 74.33        | <b>65.42</b> | <b>48.33</b> | <b>60.42</b> | 47.22        | 85.74        | <b>93.16</b> | 60.50        |
| <b>Planner + Generator</b>     |              |              |              |              |              |              |              |              |
| RePlan + Qwen-2511             | 70.33        | 58.75        | 52.92        | 59.58        | 49.44        | 84.62        | 90.41        | 59.25        |
| PromptEnhancerV2 + Qwen-2511   | 76.33        | 68.75        | 60.83        | 66.67        | 56.11        | 88.62        | 91.26        | 66.75        |
| ThinkRL-Edit (DARS config.)    | 76.00        | 64.58        | 65.42        | 70.42        | 57.78        | 89.06        | 91.84        | 67.75        |
| EditThinker + Qwen-2511        | <b>77.33</b> | 67.08        | 63.33        | 70.00        | 60.56        | 89.48        | 92.18        | 68.50        |
| DDA-Thinker-32B + Qwen-2511    | <b>79.00</b> | 71.25        | 70.83        | 75.42        | 64.44        | 91.17        | 93.10        | 72.92        |
| DARS                           | <b>79.00</b> | <b>72.50</b> | <b>71.67</b> | <b>76.25</b> | <b>66.11</b> | <b>91.36</b> | <b>93.18</b> | <b>73.75</b> |
| <b>Unified MLLMs</b>           |              |              |              |              |              |              |              |              |
| UniREdit-BAGEL                 | 72.00        | 57.08        | 55.00        | 62.92        | 47.78        | 85.68        | 89.34        | 60.17        |
| UniReason 1.0                  | 74.00        | 62.50        | 60.42        | 67.92        | 53.33        | 87.81        | 90.26        | 64.67        |
| BAGEL                          | 64.00        | 43.75        | 40.00        | 51.67        | 36.67        | 80.80        | 86.72        | 48.58        |
| BAGEL + DuplexBridge           | 79.00        | 70.83        | 74.17        | <b>77.08</b> | <b>67.78</b> | 91.27        | 92.74        | 74.33        |
| Qwen-2511                      | 71.67        | 56.25        | 51.25        | 59.17        | 40.00        | 84.30        | 90.86        | 57.25        |
| Qwen-2511 + DuplexBridge       | <b>82.00</b> | <b>78.33</b> | <b>78.33</b> | <b>80.83</b> | <b>73.33</b> | <b>93.16</b> | <b>94.21</b> | <b>79.00</b> |
| Wan-Image-7B                   | 68.33        | 52.08        | 46.67        | 56.67        | 37.78        | 82.70        | 88.47        | 53.83        |
| Wan-Image-7B + DuplexBridge    | <b>85.67</b> | <b>79.17</b> | 77.08        | 67.08        | <b>73.33</b> | 92.31        | 93.86        | 77.08        |

**Table 3 Public editing benchmarks.** Higher is better; metric scales follow section 5. Dark/light pink marks the best/second-best value per column among displayed systems.

| Method           | KRIS ↑       | RISE ↑       | ImgEdit ↑   | GEEdit EN ↑ |
|------------------|--------------|--------------|-------------|-------------|
| PromptEnhancerV2 | 71.54        | 18.90        | 4.18        | 7.59        |
| DARS             | <b>80.72</b> | 27.50        | 4.39        | <b>7.86</b> |
| BAGEL            | 56.21        | 5.80         | 3.45        | 6.40        |
| + DuplexBridge   | 76.48        | 26.11        | 4.04        | 7.41        |
| Qwen-2511        | 64.42        | 17.50        | 4.36        | 6.97        |
| + DuplexBridge   | <b>82.16</b> | <b>31.11</b> | <b>4.43</b> | <b>7.89</b> |
| Wan-Image-7B     | 60.84        | 13.06        | 4.09        | 6.88        |
| + DuplexBridge   | 80.13        | 29.17        | 4.36        | 7.74        |

**Table 4** Native UMM configurations on RISEBench-360. Qwen-2511 denotes Qwen-Image-Edit-2511. Overall is task-weighted across 85 temporal, 90 causal, 100 spatial, and 85 logical items. Dark/light pink mark the best/second-best value in each column, including ties.

| Method                      | Temporal     | Causal       | Spatial      | Logical      | Overall      |
|-----------------------------|--------------|--------------|--------------|--------------|--------------|
| BAGEL (B0)                  | 14.12        | 23.33        | 23.00        | 5.88         | 16.94        |
| BAGEL + DuplexBridge        | 24.71        | 33.33        | <b>31.00</b> | 14.12        | 26.11        |
| Qwen-2511 (B0)              | 21.18        | 26.67        | 29.00        | 9.41         | 21.94        |
| Qwen-2511 + DuplexBridge    | <b>31.76</b> | <b>38.89</b> | <b>34.00</b> | <b>18.82</b> | <b>31.11</b> |
| Wan-Image-7B (B0)           | 16.47        | 23.33        | 24.00        | 7.06         | 18.06        |
| Wan-Image-7B + DuplexBridge | 29.41        | 35.56        | <b>34.00</b> | 16.47        | 29.17        |

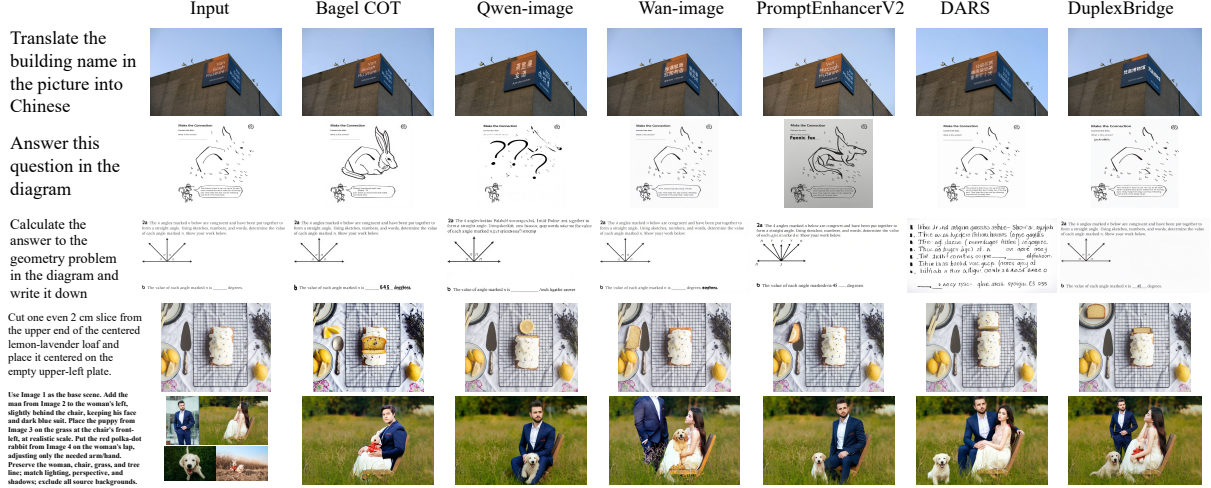


**Figure 4 Qualitative comparison on CE-Bench.** Rows cover sign translation, image-grounded question answering, geometric reasoning, spatially specified slicing, and composition from multiple references. The geometry example requires an answer within the existing diagram; the final row combines entity placement with preservation of the base scene.

respective  $B_0$  (table 4). Warm-start improvements are not uniform: Qwen preservation falls from 90.86 before adaptation to 90.14 at  $B_0$ , then reaches 94.21 with DuplexBridge. Its ImgEdit score likewise falls from 4.36 before adaptation to 4.28 at  $B_0$ , then reaches 4.43 with DuplexBridge.

**Table 5 Supervised initialization controls.**  $B_0$  uses native PE and structured SFT, before RL. A–E follow table 1; public metrics follow section 5.4. Dark/light pink mark the best/second-best values, including ties.

| CE-Bench          |              |              |              |              |              |              |              |              |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Model             | A            | B            | C            | D            | E            | AC           | Pres.        | CCR          |
| BAGEL (B0)        | 68.67        | 49.58        | 47.92        | 62.92        | 47.78        | 83.91        | 88.63        | 56.42        |
| Qwen-2511 (B0)    | <b>73.00</b> | <b>61.25</b> | <b>55.00</b> | <b>64.58</b> | <b>51.11</b> | <b>86.62</b> | <b>90.14</b> | <b>62.08</b> |
| Wan-Image-7B (B0) | 70.00        | 57.92        | 52.50        | 62.08        | 49.44        | 85.28        | 89.21        | 59.42        |

| Public benchmarks |              |              |             |             |
|-------------------|--------------|--------------|-------------|-------------|
| Model             | KRIS         | RISE         | ImgEdit     | GEEdit EN   |
| BAGEL (B0)        | 64.23        | 16.94        | 3.74        | 6.93        |
| Qwen-2511 (B0)    | <b>70.36</b> | <b>21.94</b> | <b>4.28</b> | <b>7.28</b> |
| Wan-Image-7B (B0) | 66.82        | 18.06        | 4.16        | 7.09        |

### 6.3 Information-matched communication and verification

**Table 6** Wan-Image interface ablations at SFT. Rows compare prose, flat and grouped structures, and CE-v1 variants without the redundant goal or with shuffled fields. AC and CCR are percentages.

| Interface                    | CE AC | CE CCR |
|------------------------------|-------|--------|
| Fact-matched prose           | 84.38 | 57.00  |
| Flat key-value               | 84.16 | 56.67  |
| Flat JSON                    | 84.55 | 57.42  |
| Grouped key-value            | 85.09 | 59.00  |
| CE-v1 grouped JSON           | 85.28 | 59.42  |
| CE-v1 without redundant goal | 85.17 | 59.17  |
| CE-v1 shuffled fields        | 85.02 | 58.92  |

With task facts, training data, budget, and prediction source matched, grouping raises CCR from 56.67 to 59.00 for key-value serialization and from 57.42 to 59.42 for JSON (table 6). The smaller

difference between grouped syntaxes suggests that organization matters more than JSON syntax. Removing an information-redundant goal or shuffling fields changes CCR little. These SFT-stage endpoints do not isolate token-length or truncation effects; [table 11](#) separately tests structure during gated RL.

**Table 7** Wan-Image execution-check construction. Fidelity, task recall, and duplicates use all rubric attempts. Human agreement is conditional on an adjudicated content-matched subset. The final column is CCR after separate retraining, not fixed-output rescoring.

| Construction                        | Fidelity | Recall | Dup. | Human agr. | Retrained CCR |
|-------------------------------------|----------|--------|------|------------|---------------|
| Same-contract prose rubric          | 95.11    | 88.73  | 4.28 | 87.14      | 74.92         |
| Same-contract atomic reconstruction | 95.46    | 89.12  | 3.17 | 87.36      | 75.58         |
| Compiled execution checks           | 98.21    | 91.68  | 0.73 | 87.92      | 77.08         |

The construction comparison retains the same received contract, global evaluator, and prompt checks ([table 7](#)). Source fidelity measures correct target, scope, state, and polarity relative to that contract; original-task recall uses the task inventory, and duplication measures repeated checks. These diagnostics cover all construction attempts, while human agreement uses an adjudicated content-matched subset. Compiling checks improves source fidelity and reduces duplication, but does not guarantee correct visual judgments. The CCR column comes from separate retraining with each construction method; it differs from both rescoring fixed outputs and changing G-local’s reference from the received condition to the original task ([table 9](#)).

## 6.4 Joint adaptation and supervision references

**Table 8** Wan-Image-7B training-direction ablation. All RL variants start at  $B_0$  with common supervision and reference access. Native PE omits CE-v1 SFT; U-only and G-only are trained independently. UT/DARS-style rows transplant mechanisms into Wan-Image. Dashes in the local-only row denote unreported metrics. GEdit EN uses Full-set  $G_O$ .

| Variant              | U/G RL | CE AC | CE CCR | Pres. | KRIS  | RISE  | ImgEdit | GEdit EN |
|----------------------|--------|-------|--------|-------|-------|-------|---------|----------|
| Direct, no PE        | –/–    | 82.70 | 53.83  | 88.47 | 60.84 | 13.06 | 4.09    | 6.88     |
| Native PE            | –/–    | 83.86 | 56.67  | 88.92 | 63.41 | 15.83 | 4.11    | 6.97     |
| $B_0$ : matched SFT  | –/–    | 85.28 | 59.42  | 89.21 | 66.82 | 18.06 | 4.16    | 7.09     |
| U-only RL            | ✓/–    | 87.96 | 66.67  | 88.74 | 73.28 | 24.44 | 4.17    | 7.21     |
| G-only RL            | –/✓    | 88.61 | 67.50  | 92.28 | 71.94 | 21.94 | 4.31    | 7.52     |
| Dual-role local-only | ✓/✓    | –     | 68.25  | –     | –     | –     | –       | –        |
| Joint RL, ungated    | ✓/✓    | 89.72 | 70.92  | 91.82 | 76.27 | 26.11 | 4.34    | 7.58     |
| UT-style staged      | ✓/✓    | 90.38 | 72.67  | 92.14 | 77.39 | 26.94 | 4.32    | 7.60     |
| DARS-style routing   | ✓/✓    | 90.94 | 74.08  | 92.57 | 78.68 | 27.78 | 4.37    | 7.71     |
| DuplexBridge, full   | ✓/✓    | 92.31 | 77.08  | 93.86 | 80.13 | 29.17 | 4.36    | 7.74     |

Joint ungated training reaches 70.92 CCR, above independently trained U-only (66.67) and G-only (67.50) variants ([table 8](#)). Dual-role local-only training retains both valid local losses and omits global losses and their KL terms; adding selected global feedback raises CCR from 68.25 to 77.08. All RL variants are separate runs from  $B_0$ , rather than sequential phases.

UT-style staged and DARS-style routing transplant mechanisms into Wan-Image, reaching 72.67 and 74.08 CCR. They complement the complete-system comparisons in [table 3](#). Full is slightly below the DARS-style transplant on ImgEdit (4.36 versus 4.37) and above it on GEdit EN (7.74 versus 7.71), so complete-compliance gains do not imply gains on every native metric.

With both gates retained, G-local supervision referenced to the condition actually received reaches 77.08 CCR, versus 75.00 when it instead checks the original task ([table 9](#)). U-local and global feedback retain the original task reference. This comparison isolates the specified local reference within the matched setting.

**Table 9 Additional mechanism endpoints on Wan-Image.** CCR is a percentage; the final column gives full minus control (percentage points). Count-matched and group-structure-matched random gates are separate controls. The G-local comparison changes the reference for local supervision from the original task to the received condition. These endpoints compare training variants.

| Condition                                         | CCR $\uparrow$ | Full minus control |
|---------------------------------------------------|----------------|--------------------|
| Count-matched random gate                         | 71.33          | 5.75               |
| Group-structure-matched random gate               | 72.25          | 4.83               |
| Original-task G-local source                      | 75.00          | 2.08               |
| Full: semantic gates + received-condition G-local | 77.08          | —                  |

## 6.5 Selecting reliable global feedback

**Table 10** Wan-Image global-reward gate ablations. The first four rows vary the two semantic trust decisions, with local supervision and reward-validity conditions retained. Oracle gates use extra annotations.

| Gate                             | CE CCR | KRIS  | RISE  |
|----------------------------------|--------|-------|-------|
| Neither global gate              | 70.92  | 76.27 | 26.11 |
| Prompt trust only                | 74.17  | 78.62 | 28.06 |
| Execution trust only             | 73.33  | 78.14 | 27.50 |
| Both global gates                | 77.08  | 80.13 | 29.17 |
| Count-matched random             | 71.33  | 76.48 | 26.11 |
| Soft semantic weighting          | 76.33  | 79.76 | 28.89 |
| Annotation-assisted oracle gates | 78.67  | 81.07 | 30.00 |

Both prompt and execution trust contribute: either gate alone reaches 74.17 or 73.33 CCR, compared with 77.08 for both and 70.92 with neither (table 10). Gate ablations bypass semantic trust while preserving judgment-confidence and availability tests and both local losses. Soft semantic weighting reaches 76.33. Annotation-assisted oracle gates reach 78.67 CCR, 81.07 KRIS, and 30.00 RISE; they use extra adjudicated annotations and are diagnostics rather than a guaranteed upper bound.

Semantic gating exceeds group-structure-matched random selection by 4.83 points (77.08 versus 72.25). Count-matched random selection reaches 71.33. A count-matched image mask can fragment prompt groups, change eligible group sizes, and skip more components; for U, image selection also determines whether a candidate has any accepted image. Group matching constrains these differences more closely, but neither control uniquely separates selection quality from effective optimizer exposure.

**Table 11** Structure  $\times$  global gating on Wan-Image. CE CCR (%) with matched task facts, local checks, initialization, data, and training budget.

| Interface                 | Ungated | Gated |
|---------------------------|---------|-------|
| Flat                      | 67.42   | 70.75 |
| Structured                | 70.92   | 77.08 |
| Interaction: +2.83 points |         |       |

Flat and structured conditions retain the same task facts, local checks, initialization, data, and training budget; grouping is the varied factor. With  $(r_{00}, r_{01}, r_{10}, r_{11}) = (67.42, 70.75, 70.92, 77.08)$  for structure  $s \in \{0, 1\}$  and gating  $g \in \{0, 1\}$ , the endpoint interaction is

$$\Delta_{SG} = (r_{11} - r_{10}) - (r_{01} - r_{00}) = 6.16 - 3.33 = 2.83 \text{ percentage points.} \quad (15)$$

This summarizes complementary endpoint effects without identifying a causal pathway through verifier quality.

**Table 12** Global-feedback selection on one frozen pool of 6,000 rollouts. Coverage uses all rollouts; errors use the corresponding global-eligible subset, not all module training. Wrong-prompt and unexecuted-prompt labels are assigned using an external VLM together with human annotation. These fixed-pool diagnostics differ from retrained endpoint scores.

| Gate                  | G-global cov. | Wrong / G-global | U-global cov. | Unexec. / U-global |
|-----------------------|---------------|------------------|---------------|--------------------|
| Neither semantic gate | 93.60         | 18.59            | 91.40         | 27.41              |
| Prompt trust only     | 71.20         | 2.90             | 91.40         | 27.41              |
| Execution trust only  | 93.60         | 18.59            | 64.80         | 4.60               |
| Both trust gates      | 71.20         | 2.90             | 64.80         | 4.60               |
| Count-matched random  | 71.20         | 18.12            | 64.80         | 27.11              |

On the same 6,000 fixed rollouts, the wrong-prompt rate among eligible G-global samples falls from 18.59% to 2.90%, and the unexecuted-prompt rate among eligible U-global samples falls from 27.41% to 4.60%. Labels combine external-VLM judgments and human annotation. Coverage divides accepted counts by all 6,000 rollouts; error rates divide by the respective accepted counts. These conditional rates describe selection on fixed outputs, whereas the retrained endpoints measure editing performance. They are not module-wide training fractions.

## 6.6 Checkpoint cross-pairing

**Table 13** Wan-Image checkpoint cross-pairing. U0/G0 are B0; Ut/Gt come from the same full training run, not the separately trained U-only/G-only conditions.

| Pair  | CE CCR | KRIS  | RISE  |
|-------|--------|-------|-------|
| U0 G0 | 59.42  | 66.82 | 18.06 |
| Ut G0 | 65.92  | 72.48 | 22.22 |
| U0 Gt | 68.83  | 73.94 | 24.17 |
| Ut Gt | 77.08  | 80.13 | 29.17 |

U0/G0 are the shared B0 states; Ut/Gt are checkpoints from the same full run. They differ from the independently trained U-only/G-only conditions. The CCR difference-in-differences is

$$77.08 - 65.92 - 68.83 + 59.42 = 1.75 \text{ percentage points.} \quad (16)$$

It describes checkpoint-specific partner dependence. KRIS and RISE also improve under the learned pair, as reported in [table 13](#). The cross-pair statistic does not isolate architecture sharing or provide a causal decomposition.

## 6.7 Requirement load and joint fulfillment

We compare ungated and full Wan-Image outputs on all 1,200 CE-Bench tasks, paired by task. Public load  $K$  counts change and preservation groups from the original task, independently of the predicted bridge; it differs from the number of requested edits. Let  $C$  and  $P$  indicate that all change or all preservation groups pass, respectively. Complete success is  $CP$ , while Pres. in [equation \(14\)](#) averages preservation-group pass rates within each task.

Full has higher CCR point estimates at every observed load, with an overall paired gain of 6.17 points [4.50, 7.83] ([table 14](#)). Gains are not monotonic: the  $K = 9$  contrast is 3.33 points [−0.56, 7.22], leaving its direction uncertain. At this load, change all-pass falls from 73.33% to 60.56% (−12.78 points [−17.78, −7.78]), while preservation all-pass rises from 49.44% to 52.78%. Higher joint success therefore coexists with worse change execution.

Across all tasks, change all-pass rises from 80.42% to 83.50%, preservation all-pass from 76.75% to 81.25%, and mean failed groups fall from 0.659 to 0.508. The joint states appear in [figure 5](#). Task-paired transitions include 92 joint failures becoming successes and 18 successes becoming failures; 73 tasks gain preservation all-pass and 19 lose it. At  $K = 7$ , preservation-only failures

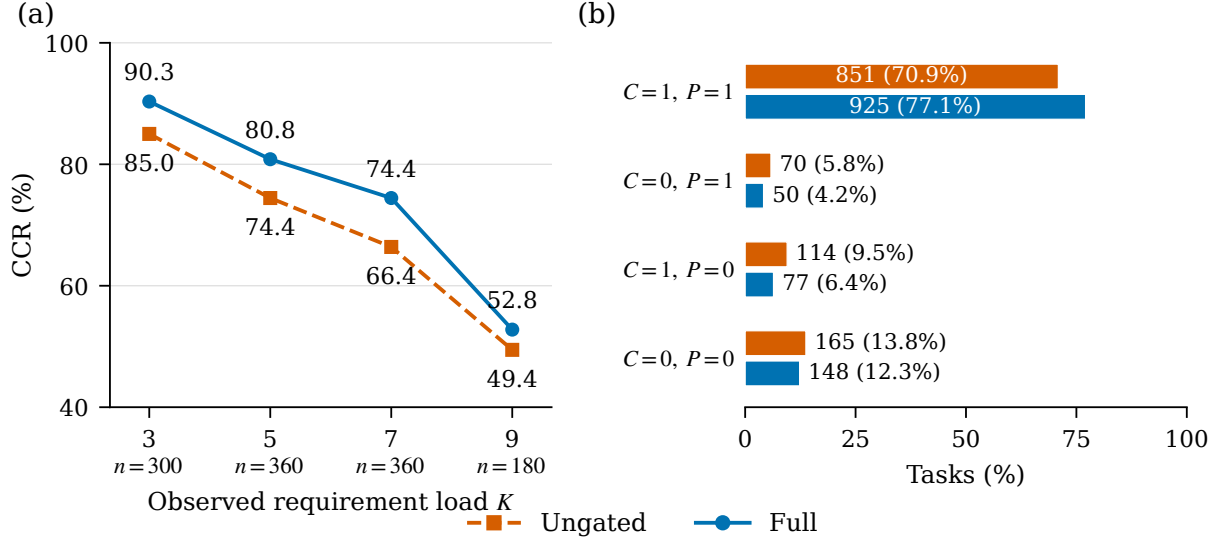


**Figure 5 Requirement load and joint change-preservation outcomes.** (a) CCR for full and ungated on the same 1,200 public tasks, grouped by their observed total number  $K$  of change and preservation requirements; each stratum’s task count is shown below its tick. Task-category composition varies across strata, so these observations do not isolate a causal effect of increasing  $K$ . (b) Counts and percentages of the same tasks in the four joint states (ungated above, full below), derived from task-level change and invariant failure counts.  $C = 1$  means every requested change passes, and  $P = 1$  means every preservation requirement passes;  $P$  is an all-pass indicator, distinct from the mean preservation fraction reported as Pres. The fraction of tasks in the  $C = P = 1$  state equals CCR.

**Table 14 Complete and average compliance by public requirement load.** U/F denote ungated/full Wan-Image; CCR and AC are percentages.  $\Delta$  is full minus ungated CCR in points, calculated before rounding. Pointwise paired bootstrap intervals follow the protocol in [section 6.7](#).

| $K$ | $N$  | U CCR | F CCR | $\Delta$ | 95% interval    | U AC  | F AC  |
|-----|------|-------|-------|----------|-----------------|-------|-------|
| 3   | 300  | 85.00 | 90.33 | +5.33    | [+2.33, +8.33]  | 95.00 | 96.78 |
| 5   | 360  | 74.44 | 80.83 | +6.39    | [+3.33, +9.44]  | 89.78 | 92.33 |
| 7   | 360  | 66.39 | 74.44 | +8.06    | [+4.72, +11.39] | 88.53 | 92.58 |
| 9   | 180  | 49.44 | 52.78 | +3.33    | [-0.56, +7.22]  | 83.15 | 84.26 |
| All | 1200 | 70.92 | 77.08 | +6.17    | [+4.50, +7.83]  | 89.72 | 92.31 |

increase from 16.11% to 17.50%; at  $K = 9$ , both-component failures increase from 26.67% to 39.44%. Aggregate improvements therefore do not extend to every component or load.

Load strata also differ in category composition:  $K = 3$  has no D tasks, while  $K = 9$  has no A tasks and includes 95/180 D tasks. Category C at  $K = 9$  has a  $-6.67$ -point CCR contrast on 30 tasks. These observed strata do not identify a causal complexity effect; requiring every group to pass can itself lower CCR as the group count grows.

Intervals use 50,000 paired task bootstrap resamples, preserving category counts overall and within each reported load. The 2.5th and 97.5th percentiles give pointwise intervals conditional on the observed strata, outputs, and labels; contrasts are computed before rounding. The human evaluation below uses a separate matched population.

## 6.8 Reward quality, cost, and downstream effects

We compare base direct scoring (M0), explicit thinking (M1), preference SFT (M2), decision-reason training with direct deployment (M3), hard judgment KD (M4), and soft judgment KD (M5). The main experiments use M5.

**Judgment quality and reliability.** Table 15 reports HPA, difficult-pair accuracy, NLL, Brier score, SwapErr, and A/B gap. The binary metric specification uses  $s = P_{\text{Yes}}/(P_{\text{Yes}} + P_{\text{No}})$ :

$$\begin{aligned} \text{NLL} &= -\frac{1}{N} \sum_n [y_n \log s_n + (1 - y_n) \log(1 - s_n)], \\ \text{Brier} &= \frac{1}{N} \sum_n (s_n - y_n)^2, \quad \text{SwapErr} = \frac{1}{N} \sum_n |s(x_n) + s(x_n^{\text{swap}}) - 1|. \end{aligned} \quad (17)$$

HPA measures agreement with human preference. The A/B gap is  $100|\text{Acc}_A - \text{Acc}_B|$  percentage points, comparing accuracy with the preferred candidate in the two input positions. NLL, Brier, and SwapErr characterize probability quality and sensitivity to candidate order.

**Table 15 Reward-model judgment quality and cost.** HPA/Hard are accuracies (%); A/B gap is in percentage points. NLL/Brier/SwapErr follow equation (17). Latency is one end-to-end pairwise call relative to M0; only M1 reasons online.

| RM                  | HPA $\uparrow$ | Hard $\uparrow$ | NLL $\downarrow$ | Brier $\downarrow$ | SwapErr $\downarrow$ | A/B gap $\downarrow$ | Latency $\downarrow$ |
|---------------------|----------------|-----------------|------------------|--------------------|----------------------|----------------------|----------------------|
| M0: Base direct     | 77.6           | 66.5            | 0.489            | 0.158              | 0.058                | 2.2                  | 1.0×                 |
| M1: Explicit think  | 82.3           | 75.8            | 0.377            | 0.120              | 0.045                | 1.3                  | 6.0×                 |
| M2: Preference SFT  | 78.9           | 69.2            | 0.439            | 0.142              | 0.051                | 1.0                  | 1.0×                 |
| M3: Decision-reason | 79.9           | 71.2            | 0.409            | 0.133              | 0.046                | 0.8                  | 1.0×                 |
| M4: Hard KD         | 80.4           | 72.3            | 0.501            | 0.154              | 0.062                | 1.6                  | 1.0×                 |
| M5: Soft KD         | 81.7           | 74.8            | 0.385            | 0.122              | 0.041                | 0.6                  | 1.0×                 |

**Table 16 Difficult-pair accuracy (%).** Scores are shown for five error categories and the reported overall.

| Category                  | M0   | M1   | M4   | M5   |
|---------------------------|------|------|------|------|
| Wrong target or ordinal   | 70.0 | 78.3 | 75.8 | 78.3 |
| Near-miss text            | 60.0 | 72.5 | 68.3 | 70.8 |
| Preservation damage       | 69.2 | 76.7 | 73.3 | 78.3 |
| Coupled-edit omission     | 61.7 | 73.3 | 68.3 | 70.8 |
| Visually good, task wrong | 71.7 | 78.3 | 75.8 | 75.8 |
| Reported overall          | 66.5 | 75.8 | 72.3 | 74.8 |

Soft KD improves difficult-pair accuracy from 66.5% to 74.8%, approaching explicit thinking’s 75.8% at 1.0× rather than 6.0× online latency (table 15). Gains over base direct span all five difficult-pair categories (table 16); M5 matches M1 on target/ordinal pairs and exceeds it on preservation damage, while M1 leads on the other three categories.

**External reward benchmarks.** Table 17 keeps MMRB2 image-editing pair accuracy (Hu et al., 2025) separate from EditReward-Bench candidate-count-dependent ranking scores (Wu et al., 2026). M5 improves MMRB2 from 66.2 to 70.6, and EditReward  $K = 2, 3, 4$  from 57.2/35.9/10.8 to 61.0/43.5/15.8. M1 remains higher on all four reported external metrics.

**Downstream reward-gate interaction.** Table 18 reports all five configurations. D0–D3 cross base/distilled direct reward with ungated/gated learning; D4 uses explicit thinking with gates. Relative to base direct, M5 raises CCR by 1.84 points without gates and 2.83 with gates. Gating adds 5.17 points under M0 and 6.16 under M5; the descriptive interaction is  $6.16 - 5.17 = 0.99$  points. This reward-gate interaction differs from the structure-gate interaction of 2.83 points in section 6.5. M1 with gates reaches 77.50 CCR, 0.42 above M5, while M5 has slightly higher preservation (93.86 versus 93.73). D0–D4 use equal update steps and rollout budgets.

**Table 17 External RM evaluation.** MMRB2 editing pair accuracy and EditReward-Bench  $K = 2, 3, 4$  ranking scores are reported separately (%).

| RM | MMRB2 edit pair acc. $\uparrow$ | EditReward $K = 2 \uparrow$ | EditReward $K = 3 \uparrow$ | EditReward $K = 4 \uparrow$ |
|----|---------------------------------|-----------------------------|-----------------------------|-----------------------------|
| M0 | 66.2                            | 57.2                        | 35.9                        | 10.8                        |
| M1 | 71.8                            | 61.8                        | 44.9                        | 16.6                        |
| M2 | 67.1                            | 58.3                        | 37.8                        | 11.9                        |
| M3 | 68.6                            | 59.6                        | 40.4                        | 13.3                        |
| M4 | 69.3                            | 60.1                        | 41.5                        | 13.8                        |
| M5 | 70.6                            | 61.0                        | 43.5                        | 15.8                        |

**Table 18 Reward  $\times$  gate comparisons on Wan-Image.** D0–D3 form the  $2 \times 2$  comparison; D4 uses explicit thinking. D0–D4 use equal update steps and rollout budgets. All metrics are percentages.

| ID / RM | Gates | CCR $\uparrow$ | AC $\uparrow$ | Pres. $\uparrow$ |
|---------|-------|----------------|---------------|------------------|
| D0 / M0 | Off   | 69.08          | 88.91         | 91.35            |
| D1 / M0 | On    | 74.25          | 91.28         | 93.12            |
| D2 / M5 | Off   | 70.92          | 89.72         | 91.82            |
| D3 / M5 | On    | 77.08          | 92.31         | 93.86            |
| D4 / M1 | On    | 77.50          | 92.42         | 93.73            |

## 6.9 Human evaluation of final edits

The evaluation contains 200 CE-Bench tasks, with 50/40/40/40/30 tasks in categories A–E. Direct, ungated, and full have 5, 4, and 3 unresolved human success labels, respectively; automatic labels are available for all outputs. Per-method CCR, bias (automatic minus human CCR), and agreement use each method’s human-decidable subset. Table 19 also reports false acceptance/rejection and bounds over all 200 tasks that vary the unresolved labels. These final-output judgments are distinct from execution-check construction agreement in table 7.

**Table 19 Final-output human evaluation on 200 CE-Bench tasks.**  $N$  is the number of human-decidable outputs per method. CCR, agreement, false acceptance (FA), and false rejection (FR) are percentages; bias is automatic minus human CCR in points. FA conditions on human failures and FR on human successes. Bounds assign unresolved labels to failure or success over all 200 tasks; they are identification bounds, not confidence intervals.

| Method  | $N$ | Human CCR | Auto. CCR | Bias  | Agree. | FA    | FR   | Human CCR bounds |
|---------|-----|-----------|-----------|-------|--------|-------|------|------------------|
| Direct  | 195 | 53.33     | 54.36     | +1.03 | 88.72  | 13.19 | 9.62 | [52.0, 54.5]     |
| Ungated | 196 | 69.39     | 70.92     | +1.53 | 91.33  | 16.67 | 5.15 | [68.0, 70.0]     |
| Full    | 197 | 76.14     | 76.65     | +0.51 | 93.40  | 14.89 | 4.00 | [75.0, 76.5]     |

**Matched gains and scoring bias.** Full and ungated share 194 tasks with both human labels resolved. On this common population, human CCR rises from 69.07% to 75.77%, a 6.70-point gain [1.03, 12.37] (table 20). The automatic gain is 6.19 points; automatic minus human gain is  $-0.52$  points  $[-3.61, 2.58]$ , leaving the direction of gain bias inconclusive. Full’s per-method CCR of 76.14% uses 197 decidable outputs and therefore differs from the common-population rate.

**Dimension-specific preference.** The same 200 tasks compare full against ungated and direct on instruction fulfillment, preservation, and visual quality (table 21). Net preference is  $100(N_{\text{win}} - N_{\text{loss}})/N_{\text{decidable}}$ , with ties included and unresolved judgments counted separately. Instruction preference favors full in both comparisons. Against ungated, preservation and quality intervals span zero; against direct, preservation also favors full, while the quality interval includes zero. Figure 6 visualizes the ungated comparison. These evaluations share a task pool; failure to detect

**Table 20 Full versus ungated on 194 common tasks.** Point estimates and 95% percentile intervals are in percentage points. Intervals use 50,000 paired task bootstrap resamples within categories, preserving the observed category counts of this common decidable set.

| Quantity                        | Estimate | 95% interval    |
|---------------------------------|----------|-----------------|
| Human gain $\Delta_H$           | +6.70    | [+1.03, +12.37] |
| Automatic gain $\Delta_A$       | +6.19    | [+1.55, +10.82] |
| Gain bias $\Delta_A - \Delta_H$ | -0.52    | [-3.61, +2.58]  |

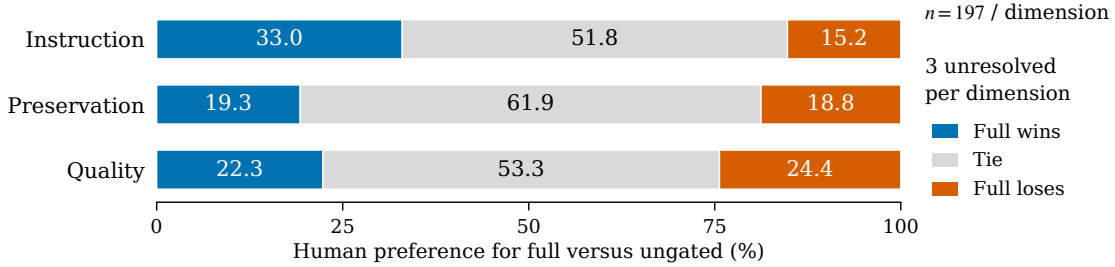


**Figure 6 Human preferences for full versus ungated.** Each dimension has 197 decidable task pairs and three unresolved pairs out of 200. Wins, ties, and losses retain their separate categories; interval estimates for net preference appear in [table 21](#).

a quality difference does not establish non-inferiority.

**Table 21 Human preference for full.** W/T/L are wins/ties/losses; Unres. counts unresolved judgments out of 200 task pairs per comparison. Net preference is  $100(W-L)/(W+T+L)$ , in points. The 95% intervals use category-stratified task bootstrap; ties remain in the denominator.

| Full comparison | Dimension    | W/T/L     | Unres. | Net    | 95% interval     |
|-----------------|--------------|-----------|--------|--------|------------------|
| vs. ungated     | Instruction  | 65/102/30 | 3      | +17.77 | [+8.63, +26.90]  |
| vs. ungated     | Preservation | 38/122/37 | 3      | +0.51  | [-8.12, +9.14]   |
| vs. ungated     | Quality      | 44/105/48 | 3      | -2.03  | [-11.68, +7.11]  |
| vs. direct      | Instruction  | 89/83/25  | 3      | +32.49 | [+22.84, +42.13] |
| vs. direct      | Preservation | 70/94/33  | 3      | +18.78 | [+9.14, +28.43]  |
| vs. direct      | Quality      | 61/87/48  | 4      | +6.63  | [-3.57, +16.84]  |

**Uncertainty.** We use 50,000 category-stratified task bootstrap resamples and percentile 95% intervals. Paired method and evaluator labels stay together; each analysis retains its category-specific decidable counts (49/39/39/39/28 for the common full-ungated population). Preference intervals use the relevant dimension’s decidable pairs. Intervals quantify task-sampling uncertainty conditional on the evaluated outputs and observed labels.

## 6.10 Text-to-image adaptation

DuplexBridge-T2I retrains both roles: U specifies requirements from text and G synthesizes from that condition. The Qwen T2I backbone is Qwen-Image; BAGEL’s baseline uses native CoT. All three configurations improve on all five metrics ([table 22](#)). These results measure complete T2I adaptation, not zero-shot transfer from editing checkpoints.

GenEval2 Soft-TIFA computes per-prompt arithmetic (AM) or geometric (GM) means of soft VQA scores, then averages the resulting prompt scores ([Kamath et al., 2025](#)); GM measures continuous compositional fulfillment. R2I tests reasoning ([Chen et al., 2025](#)), WISE-V world-knowledge consistency ([Niu et al., 2026](#)), and DPG dense prompt following ([Hu et al., 2024](#)). R2I and WISE-V use the displayed  $\times 100$  scales.

**Table 22 Text-to-image adaptation after retraining.** Metrics follow [section 6.10](#); dark/light pink mark the best/second-best values across all six systems, including ties.

| Model                           | GenEval2<br>AM $\uparrow$ | GenEval2<br>GM $\uparrow$ | R2I<br>$\times 100 \uparrow$ | WISE-V<br>$\times 100 \uparrow$ | DPG $\uparrow$ |
|---------------------------------|---------------------------|---------------------------|------------------------------|---------------------------------|----------------|
| BAGEL + native CoT              | 70.9                      | 23.1                      | 48.0                         | 62.8                            | 85.1           |
| BAGEL + DuplexBridge-T2I        | 77.8                      | 31.5                      | 56.0                         | <b>70.5</b>                     | 86.8           |
| Qwen-Image                      | 80.8                      | 33.8                      | 52.0                         | 51.0                            | 88.3           |
| Qwen-Image + DuplexBridge-T2I   | <b>85.0</b>               | <b>42.0</b>               | <b>61.0</b>                  | 66.0                            | <b>89.7</b>    |
| Wan-Image-7B                    | 79.4                      | 32.2                      | 51.0                         | 54.0                            | 87.7           |
| Wan-Image-7B + DuplexBridge-T2I | 84.2                      | 40.6                      | 59.5                         | 64.5                            | 89.3           |

## 7 Qualitative Comparisons

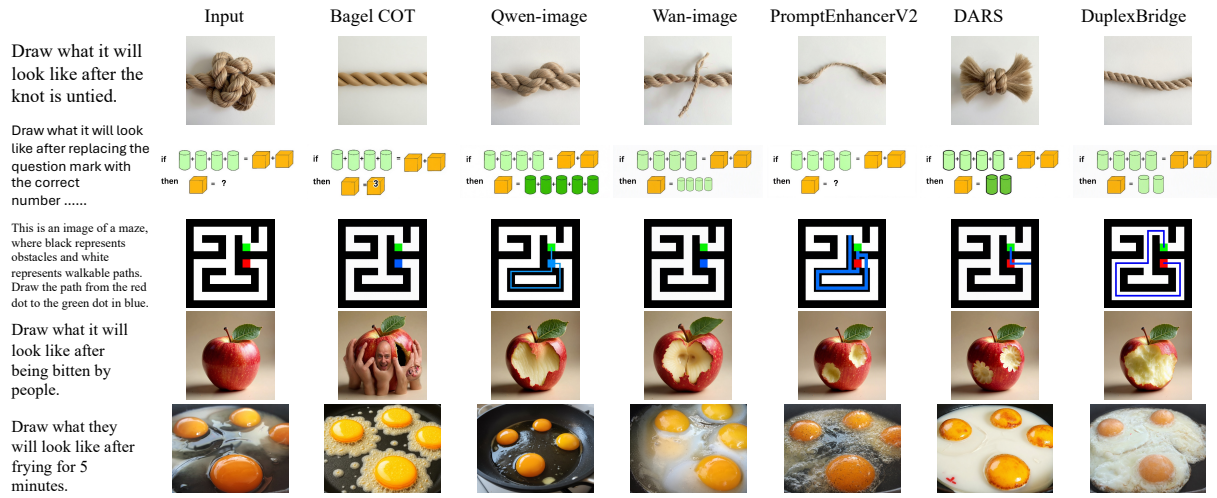


**Figure 7 Qualitative comparison on RISEBench.** Rows cover untying a knot, a visual equivalence relation, maze navigation, a bitten apple, and cooking progression. The maze requires a connected path through the white corridor; the physical transformations also require recognizable source content.

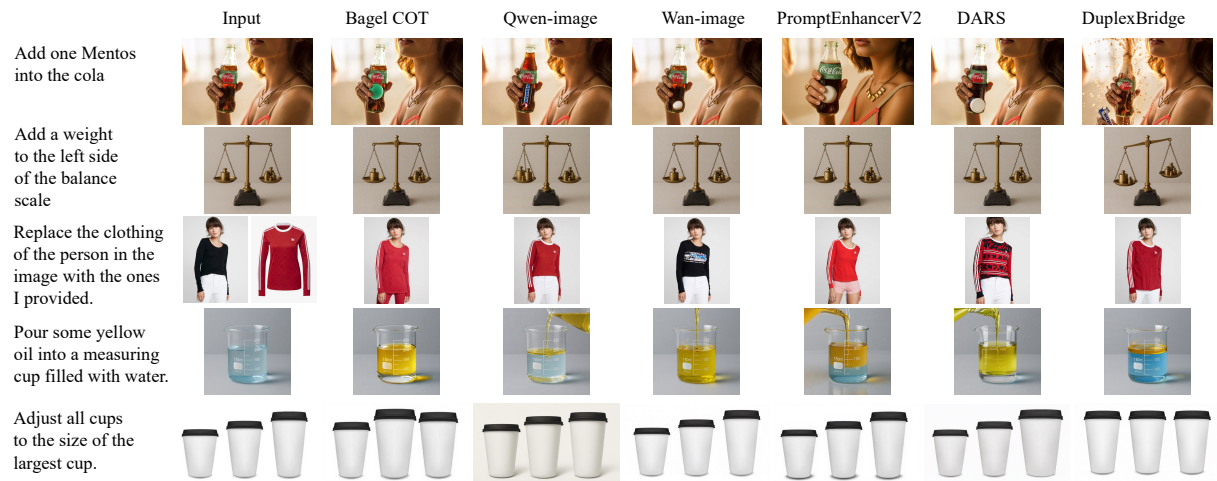


**Figure 8 Qualitative comparison on KRIS-Bench.** The tasks combine requested changes with physical consequences, reference fidelity, and preservation. Rows show Mentos–cola interaction, added load, garment replacement, oil–water separation, and cup-size equalization.



**Figure 9 Additional CE-Bench showcase.** Arrows indicate the transition from source images to edited outputs; auxiliary images provide visual references for flower replacement and dog insertion. Examples include sign-text replacement and translation, localized color changes, a traffic-light state change, and object appearance edits. The paired images show the edited regions alongside the surrounding scene.

**RISEBench.** In [figure 7](#), DuplexBridge draws a connected blue maze route through the white corridor between the red and green endpoints. The untied rope and opaque egg whites show physical state changes. In the equivalence example, DuplexBridge replaces the question mark with two green cylinders, consistent with the displayed relation and the original task’s request for the correct number of cylinders.

**KRIS-Bench.** The examples in [figure 8](#) illustrate coupled requirements: a balance responds to added load, oil separates above water, and three cups reach a common size. Reference-guided garment replacement preserves the white trousers. These examples accompany the quantitative public-benchmark results in [table 3](#).

## 8 Discussion and Limitations

DuplexBridge improves joint task fulfillment by separating local supervision references and selecting additional global feedback. Its checks remain semantic judgments: a shared condition does not guarantee correct source binding or visual assessment. Unresolved judgments abstain, and high reward confidence alone does not establish correctness.

Mean execution thresholds can also admit a failed critical requirement when other groups pass ([section 4.4](#)). The gates define selective objectives rather than counterfactual attribution or unbiased unconditional-return optimization ([section 4.4](#)). The random-selection controls narrow some group-structure differences but do not uniquely separate sample quality from optimizer exposure. Structure–gate and reward–gate interactions describe aggregate endpoints; checkpoint cross-pairing describes partner dependence within the evaluated run.

The observed requirement-load strata differ in task-category composition. Their contrasts do not identify a causal effect of complexity, and higher aggregate CCR coexists with some component regressions. Human evaluation supports an instruction-fulfillment gain on the matched population, while preservation and quality intervals against ungated include zero. These comparisons do not establish visual-quality non-inferiority. Reported bootstrap intervals quantify task-sampling uncertainty conditional on the evaluated outputs and labels, rather than variation across new training runs.

The text-to-image extension uses task-specific retraining and a redesigned interface; it is not zero-shot transfer from editing checkpoints. Our experiments have not evaluated more complex video generation or editing. Extending the method to the video domain remains a direction for future work.

## 9 Conclusion

DuplexBridge connects a unified model’s own understanding and generation through a checkable editing condition. Local supervision distinguishes task fidelity from received-condition execution, while reciprocal trust gates select additional global feedback and retain valid local learning. Across three configurations, complete compliance improves over direct editors; controlled Wan-Image comparisons support the supervision references and semantic selection, with gains also supported by human evaluation. These findings support separating local responsibilities when jointly adapting understanding and generation.

## References

- Haixiang Cao, Jiajiong Cao, Xuanpu Zhang, Changqian Yu, and Chaoqun Wang. DARS: Dual-level credit assignment RL with structured reasoning for instruction-based image editing, 2026. <https://arxiv.org/abs/2608.20161>.
- Kaijie Chen, Zihao Lin, Zhiyang Xu, Ying Shen, Yuguang Yao, Joy Rimchala, Jiaxin Zhang, and Lifu

- Huang, R2I-Bench: Benchmarking reasoning-driven text-to-image generation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 12595–12630, 2025. doi: 10.18653/v1/2025.emnlp-main.636. <https://aclanthology.org/2025.emnlp-main.636/>.
- Jaemin Cho, Yushi Hu, Jason Baldridge, Roopal Garg, Peter Anderson, Ranjay Krishna, Mohit Bansal, Jordi Pont-Tuset, and Su Wang. Davidsonian scene graph: Improving reliability in fine-grained evaluation for text-to-image generation. In *International Conference on Learning Representations*, 2024. [https://proceedings.iclr.cc/paper\\_files/paper/2024/hash/446905abfec9db8856f7f7465bbd5552-Abstract-Conference.html](https://proceedings.iclr.cc/paper_files/paper/2024/hash/446905abfec9db8856f7f7465bbd5552-Abstract-Conference.html).
- Chaorui Deng, Deyao Zhu, Kunchang Li, Chenhui Gou, Feng Li, Zeyu Wang, Shu Zhong, Weihao Yu, Xiaonan Nie, Ziang Song, Guang Shi, and Haoqi Fan. Emerging properties in unified multimodal pretraining. *arXiv preprint arXiv:2505.14683*, 2025. <https://arxiv.org/abs/2505.14683>.
- Haiwen Diao, Jiahao Wang, Chenjing Ding, Hanming Deng, Jiangnan Chen, Ruixi Zhang, Ruohui Wang, Wenwen Tong, Xiangyu Fan, Yubo Wang, Yue Zhu, Yuwei Niu, Zhengqi Bai, Zhiqian Lin, Zhitao Yang, Zhongang Cai, Bo Yang, Chen Feng, Chengguang Lv, Guangjia Liu, Guanlin Wang, Hanyu Zhang, Haojia Yu, Hongcan Xiao, Hongli Wang, Huan Wu, Huaping Zhong, Jian Fang, Jianan Fan, Jiaqi Li, Jiefan Lu, Jing Zuo, Jingcheng Ni, Junxiang Xu, Linjun Dai, Mutian Xu, Peishen Yan, Penghao Wu, Ruijie Mao, Ruisi Wang, Shihao Bai, Shuang Yang, Shuya Yang, Shuyan Zheng, Silei Wu, Siying Li, Tao Chu, Tianbo Zhong, Tongxi Zhou, Weichao Luo, Weichen Fan, Wenhao Jia, Wenjie Gao, Xiangli Kong, Yan Li, Yang Yong, Zimo Wen, Zixuan Qian, Wenxiu Sun, Ruihao Gong, Quan Wang, Lewei Lu, Lei Yang, Ziwei Liu, and Dahua Lin. SenseNova-U1.5: Towards native unified visual intelligence, 2026. <https://arxiv.org/abs/2609.11929>.
- Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11794. <https://ojs.aaai.org/index.php/AAAI/article/view/11794>.
- Tsu-Jui Fu, Wenze Hu, Xianzhi Du, William Yang Wang, Yinfei Yang, and Zhe Gan. Guiding instruction-based image editing via multimodal large language models. In *International Conference on Learning Representations*, 2024. [https://proceedings.iclr.cc/paper\\_files/paper/2024/hash/f0e91b1314fa5eabf1d7ef6d1561ecec-Abstract-Conference.html](https://proceedings.iclr.cc/paper_files/paper/2024/hash/f0e91b1314fa5eabf1d7ef6d1561ecec-Abstract-Conference.html).
- Xiwei Hu, Rui Wang, Yixiao Fang, Bin Fu, Pei Cheng, and Gang Yu. ELLA: Equip diffusion models with LLM for enhanced semantic alignment, 2024. <https://arxiv.org/abs/2403.05135>.
- Yushi Hu, Benlin Liu, Jungo Kasai, Yizhong Wang, Mari Ostendorf, Ranjay Krishna, and Noah A. Smith. TIFA: Accurate and interpretable text-to-image faithfulness evaluation with question answering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. [https://openaccess.thecvf.com/content/ICCV2023/papers/Hu\\_TIFA\\_Accurate\\_and\\_Interpretable\\_Text-to-Image\\_Faithfulness\\_Evaluation\\_with\\_Question\\_Answering\\_ICCV\\_2023\\_paper.pdf](https://openaccess.thecvf.com/content/ICCV2023/papers/Hu_TIFA_Accurate_and_Interpretable_Text-to-Image_Faithfulness_Evaluation_with_Question_Answering_ICCV_2023_paper.pdf).
- Yushi Hu, Reyhane Askari-Hemmat, Melissa Hall, Emily Dinan, Luke Zettlemoyer, and Marjan Ghazvininejad. Multimodal RewardBench 2: Evaluating Omni Reward Models for Interleaved Text and Image. *arXiv preprint arXiv:2512.16899*, 2025. doi: 10.48550/ARXIV.2512.16899. <https://arxiv.org/abs/2512.16899>.
- Zican Hu, Xuyang Hu, Yiming Liu, Zuwei Long, Wei Liu, Yunzhuo Hao, Jiawei Gu, Linjie Li, Yu Cheng, Zhenhong Sun, Weibo Gu, Xing Sun, and Zhi Wang. Bridging interleaved multi-modal reasoning as a unified decision process, 2026. <https://arxiv.org/abs/2607.03748>.
- Yuzhou Huang, Liangbin Xie, Xintao Wang, Ziyang Yuan, Xiaodong Cun, Yixiao Ge, Jiantao Zhou, Chao Dong, Rui Huang, Ruimao Zhang, and Ying Shan. SmartEdit: Exploring complex instruction-based image editing with multimodal large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8362–8371, 2024. [https://openaccess.thecvf.com/content/CVPR2024/html/Huang\\_SmartEdit\\_Exploring\\_Complex\\_Instruction-based\\_Image\\_Editing\\_with\\_Multimodal\\_Large\\_Language\\_CVPR\\_2024\\_paper.html](https://openaccess.thecvf.com/content/CVPR2024/html/Huang_SmartEdit_Exploring_Complex_Instruction-based_Image_Editing_with_Multimodal_Large_Language_CVPR_2024_paper.html).
- Zhangqi Jiang, Zheng Sun, Xianfang Zeng, Yufeng Yang, Xuanyang Zhang, Yongliang Wu, Wei Cheng, Gang Yu, Xu Yang, and Bihan Wen. Geditbench v2: A human-aligned benchmark for general image editing, 2026. <https://arxiv.org/abs/2603.28547>.
- Siyu Jiao, Yiheng Lin, Yujie Zhong, Qi She, Wei Zhou, Xiaohan Lan, Zilong Huang, Fei Yu, Yingchen Yu, Yunqing Zhao, Yao Zhao, and Yunchao Wei. ThinkGen: Generalized thinking for visual generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*,

- tion, pages 14713–14723, 2026. [https://openaccess.thecvf.com/content/CVPR2026/html/Jiao\\_ThinkGen\\_Generalized\\_Thinking\\_for\\_Visual\\_Generation\\_CVPR\\_2026\\_paper.html](https://openaccess.thecvf.com/content/CVPR2026/html/Jiao_ThinkGen_Generalized_Thinking_for_Visual_Generation_CVPR_2026_paper.html).
- Xin Jin, Huanqia Cai, Zhen Li, Zechao Zhan, Dengyang Jiang, Aiming Hao, Yuming Jiang, Xiangpeng Yang, Chunle Guo, Peng Gao, Ming-Ming Cheng, and Steven C. H. Hoi. Z-Reward: Beyond Scalar Rewards by Internalizing Reasoning into Score Distributions. *arXiv preprint arXiv:2606.09076*, 2026. doi: 10.48550/ARXIV.2606.09076. <https://arxiv.org/abs/2606.09076>.
- Amita Kamath, Kai-Wei Chang, Ranjay Krishna, Luke Zettlemoyer, Yushi Hu, and Marjan Ghazvininejad. Geneval 2: Addressing benchmark drift in text-to-image evaluation. *arXiv preprint arXiv:2512.16853*, 2025. <https://arxiv.org/abs/2512.16853>.
- Hengjia Li, Liming Jiang, Qing Yan, Yizhi Song, Hao Kang, Zichuan Liu, Xin Lu, Boxi Wu, and Deng Cai. Thinkrl-edit: Thinking in reinforcement learning for reasoning-centric image editing, 2026. <https://arxiv.org/abs/2601.03467>.
- Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-GRPO: Training flow matching models via online RL. In *Advances in Neural Information Processing Systems*, volume 38, pages 45650–45685, 2025a. doi: 10.52202/085713-1362. [https://proceedings.neurips.cc/paper\\_files/paper/2025/hash/3a10c46572628d58cb44fb705f25cbbf-Abstract-Conference.html](https://proceedings.neurips.cc/paper_files/paper/2025/hash/3a10c46572628d58cb44fb705f25cbbf-Abstract-Conference.html).
- Jie Liu, Zilyu Ye, Linxiao Yuan, Shenhan Zhu, Yu Gao, Jie Wu, Kunchang Li, Xionghui Wang, Xiaonan Nie, Weilin Huang, and Wanli Ouyang. UniGRPO: Unified policy optimization for reasoning-driven visual generation, 2026. <https://arxiv.org/abs/2603.23500>.
- Shiyu Liu, Yucheng Han, Peng Xing, Fukun Yin, Rui Wang, Wei Cheng, Jiaqi Liao, Yingming Wang, Honghao Fu, Chunrui Han, Guopeng Li, Yuang Peng, Quan Sun, Jingwei Wu, Yan Cai, Zheng Ge, Ranchen Ming, Lei Xia, Xianfang Zeng, Yibo Zhu, Binxing Jiao, Xiangyu Zhang, Gang Yu, and Daxin Jiang. Step1x-edit: A practical framework for general image editing, 2025b. <https://arxiv.org/abs/2504.17761>.
- Chaojie Mao, Chen-Wei Xie, Chongyang Zhong, Haoyou Deng, Jiaying Zhao, Jie Xiao, Jinbo Xing, Jingfeng Zhang, Jingren Zhou, Jingyi Zhang, Jun Dan, Kai Zhu, Kang Zhao, Keyu Yan, Minghui Chen, Pandeng Li, Shuangli Chen, Tong Shen, Yu Liu, Yue Jiang, Yulin Pan, Yuxiang Tuo, Zeyinzi Jiang, Zhen Han, Ang Wang, Bang Zhang, Baole Ai, Bin Wen, Boang Feng, Feiwu Yu, Gang Wang, Haiming Zhao, He Kang, Jianjing Xiang, Jianyuan Zeng, Jinkai Wang, Junjie Zhou, Ke Sun, Linqian Wu, Pei Gong, Pingyu Wu, Ruiwen Wu, Tongtong Su, Wenmeng Zhou, Wenting Shen, Wenyan Yu, Xianjun Xu, Xiaoming Huang, Xiejie Shen, Xin Xu, Yan Kou, Yangyu Lv, Yifan Zhai, Yitong Huang, Yun Zheng, Yuntao Hong, Zhe Zhang, and Zhicheng Zhang. Wan-image: Pushing the boundaries of generative visual intelligence, 2026. <https://arxiv.org/abs/2604.19858>.
- Yuwei Niu, Munan Ning, Mengren Zheng, Weiyang Jin, Bin Lin, Peng Jin, Jiaqi Liao, Chaoran Feng, Fanqing Meng, Kun-Peng Ning, Bin Zhu, and Li Yuan. WISE: World knowledge-informed semantic evaluation for text-to-image generation. In *Forty-third International Conference on Machine Learning*, 2026. <https://github.com/PKU-YuanGroup/WISE>.
- Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024. <https://arxiv.org/abs/2402.03300>.
- Linqing Wang, Zhiyong Xu, Ximing Xing, Yiji Cheng, Zhiyuan Zhao, Donghao Li, Tiankai Hang, Zhenxi Li, Jiale Tao, Qixun Wang, Ruihuang Li, Comi Chen, Xin Li, Mingrui Wu, Xincheng Deng, Shuyang Gu, Chunyu Wang, and Qinglin Lu. Promptenhancer: Taming your rewriter for text-to-image generation via fine-grained reward. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14895–14904, 2026. [https://openaccess.thecvf.com/content/CVPR2026/html/Wang\\_PromptEnhancer\\_Taming\\_Your\\_Rewriter\\_for\\_Text-to-Image\\_Generation\\_via\\_Fine-Grained\\_Reward\\_CVPR\\_2026\\_paper.html](https://openaccess.thecvf.com/content/CVPR2026/html/Wang_PromptEnhancer_Taming_Your_Rewriter_for_Text-to-Image_Generation_via_Fine-Grained_Reward_CVPR_2026_paper.html).
- Yibin Wang, Zhimin Li, Yuhang Zang, Chunyu Wang, Qinglin Lu, Cheng Jin, and Jiaqi Wang. Unified Multimodal Chain-of-Thought Reward Model through Reinforcement Fine-Tuning. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-5315. [https://proceedings.neurips.cc/paper\\_files/paper/2025/hash/e95e9f0c127aa1cfa2628adb2f3cb107-Abstract-Conference.html](https://proceedings.neurips.cc/paper_files/paper/2025/hash/e95e9f0c127aa1cfa2628adb2f3cb107-Abstract-Conference.html).
- Jie Wu, Yu Gao, Zilyu Ye, Ming Li, Liang Li, Hanzhong Guo, Jie Liu, Zeyue Xue, Xiaoxia Hou, Wei Liu, Yan Zeng, and Weilin Huang. RewardDance: Reward Scaling in Visual Generation. *arXiv preprint arXiv:2509.08826*, 2025a. doi: 10.48550/ARXIV.2509.08826. <https://arxiv.org/abs/2509.08826>.

- Keming Wu, Sicong Jiang, Max Ku, Ping Nie, Minghao Liu, and Wenhui Chen. Editreward: A human-aligned reward model for instruction-guided image editing. In *International Conference on Learning Representations*, 2026. [https://proceedings.iclr.cc/paper\\_files/paper/2026/hash/4d68af3056842229f8b356d686aec914-Abstract-Conference.html](https://proceedings.iclr.cc/paper_files/paper/2026/hash/4d68af3056842229f8b356d686aec914-Abstract-Conference.html).
- Tsung-Han Wu, Long Lian, Joseph E. Gonzalez, Boyi Li, and Trevor Darrell. Self-correcting LLM-controlled diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6327–6336, 2024. [https://openaccess.thecvf.com/content/CVPR2024/html/Wu\\_Self-correcting\\_LLM-controlled\\_Diffusion\\_Models\\_CVPR\\_2024\\_paper.html](https://openaccess.thecvf.com/content/CVPR2024/html/Wu_Self-correcting_LLM-controlled_Diffusion_Models_CVPR_2024_paper.html).
- Yongliang Wu, Zonghui Li, Xinting Hu, Xinyu Ye, Xianfang Zeng, Gang Yu, Wenbo Zhu, Bernt Schiele, Ming-Hsuan Yang, and Xu Yang. KRIS-Bench: Benchmarking next-level intelligent image editing models. In *Advances in Neural Information Processing Systems*, volume 38. Curran Associates, Inc., 2025b. doi: 10.52202/085713-5242. [https://proceedings.neurips.cc/paper\\_files/paper/2025/hash/e619b285582fb12f4c3de3a507b8b99c-Abstract-Datasets\\_and\\_Benchmarks\\_Track.html](https://proceedings.neurips.cc/paper_files/paper/2025/hash/e619b285582fb12f4c3de3a507b8b99c-Abstract-Datasets_and_Benchmarks_Track.html). Datasets and Benchmarks Track.
- Hanqing Yang, Qiang Zhou, Yongchao Du, Sashuai Zhou, Zhibin Wang, Jun Song, Tiezheng Ge, Cheng Yu, and Bo Zheng. Dda-thinker: Decoupled dual-atomic reinforcement learning for reasoning-driven image editing, 2026. <https://arxiv.org/abs/2604.25477>.
- Yang Ye, Xianyi He, Zongjian Li, Bin Lin, Shenghai Yuan, Zhiyuan Yan, Bohan Hou, and Li Yuan. Imgedit: A unified image editing dataset and benchmark. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-4387. [https://proceedings.neurips.cc/paper\\_files/paper/2025/hash/bec783f119b969772c35243b1c801090-Abstract-Datasets\\_and\\_Benchmarks\\_Track.html](https://proceedings.neurips.cc/paper_files/paper/2025/hash/bec783f119b969772c35243b1c801090-Abstract-Datasets_and_Benchmarks_Track.html). Datasets and Benchmarks Track.
- Xiangyu Zhao, Peiyuan Zhang, Kexian Tang, Xiaorong Zhu, Hao Li, Wenhao Chai, Zicheng Zhang, Renqiu Xia, Guangtao Zhai, Junchi Yan, Hua Yang, Xue Yang, and Haodong Duan. Envisioning beyond the pixels: Benchmarking reasoning-informed visual editing. In *Advances in Neural Information Processing Systems*, volume 38. Curran Associates, Inc., 2025. doi: 10.52202/085713-5791. [https://proceedings.neurips.cc/paper\\_files/paper/2025/hash/fe79898dcf078ec54b6f6eeea10ebb751-Abstract-Datasets\\_and\\_Benchmarks\\_Track.html](https://proceedings.neurips.cc/paper_files/paper/2025/hash/fe79898dcf078ec54b6f6eeea10ebb751-Abstract-Datasets_and_Benchmarks_Track.html). Datasets and Benchmarks Track.
- Dian Zheng, Manyuan Zhang, Hongyu Li, Hongbo Liu, Kai Zou, Kaituo Feng, and Hongsheng Li. Uni-Edit: Intelligent editing is a general task for unified model tuning. *arXiv preprint arXiv:2605.21487*, 2026. <https://arxiv.org/abs/2605.21487>.
- Sashuai Zhou, Qiang Zhou, Jijin Hu, Hanqing Yang, Yue Cao, Junpeng Ma, Yinchao Ma, Jun Song, Tiezheng Ge, Cheng Yu, Bo Zheng, and Zhou Zhao. Unified thinker: A general reasoning core for image generation. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10597–10613. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.484. <https://aclanthology.org/2026.acl-long.484/>.