Visual-Aware CoT: Achieving High-Fidelity Visual Consistency in Unified Models

Zixuan Ye[1], Quande Liu[2,†], Cong Wei[3], Xintao Wang[2], Yuanxing Zhang[2], Pengfei Wan[2], Kun Gai[2], Wenhan Luo[1,†]

[1]Hong Kong University of Science and Technology, [2]Kling Team, KuaiShou Technology, [3]University of Waterloo, [†]Corresponding author

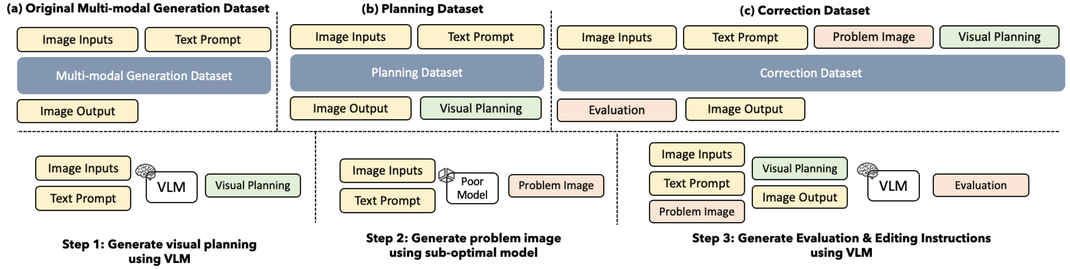
CVPR JUNE 3-7, 2026 DENVER COLORADO

Project Page

TL;DR a paradigm shift from **"text-following"** to **"visually-aware" reasoning** in generation-understanding unified models.

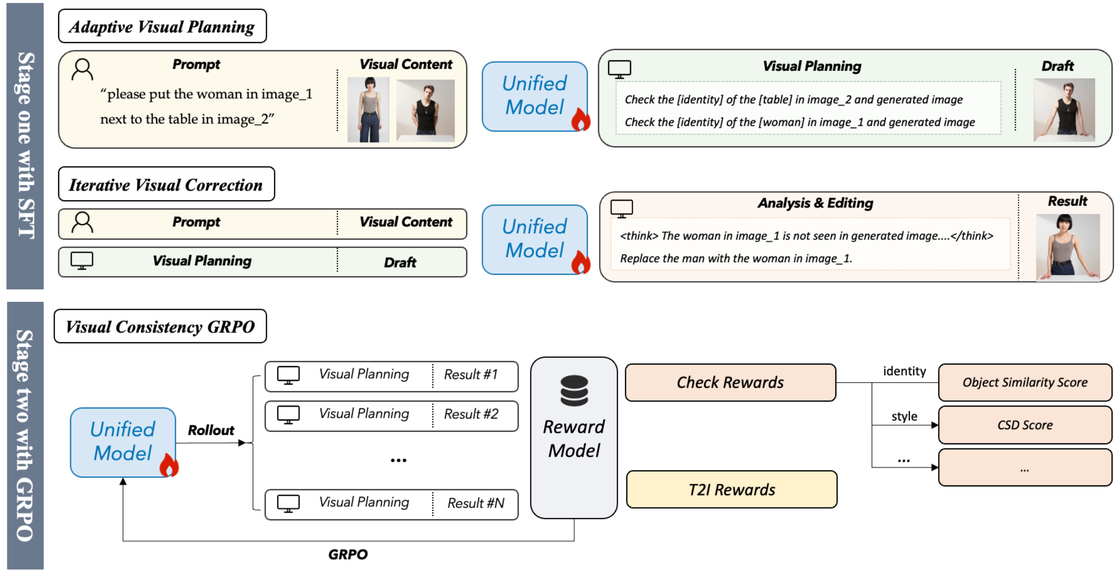
Dataset

Planning Dataset & Correction Dataset



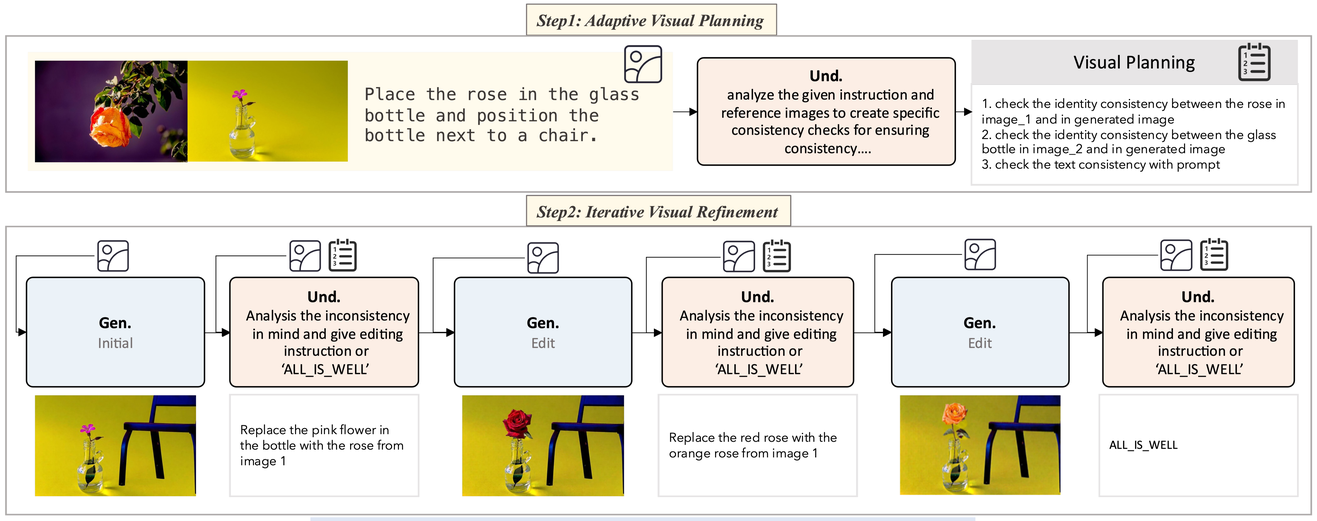
Training

SFT + GRPO



Method

Adaptive planning → Iterative Visual Refinement



Performance

Achieve High-Fidelity in Multi-reference Generation