SSR-A: Spatial- and Semantic-Aware Instructions and Curriculum Reinforcement for Advertisement Compliant Rectification

Te Cao[*] **Mengge Xue**[*†] **Zhenyu Hu** **Yuan Chen**
Liqun Liu[‡] **Peng Shu** **Huan Yu** **Jie Jiang**
Tencent
{rosaliecao,berryxue,mapleshu,izayoiychen,
liqunliu,archershu,huanyu,zeus}@tencent.com

Abstract

While advertising is a cornerstone of commercial growth, it is constrained by online violation detection systems that reject non-compliant content at a million-scale daily. Advertisers urgently require automated solutions to rectify these advertisements, especially visual ads, as manual fixing is unscalable. Although recent safety-driven methods can achieve compliance, they typically suffer from over-editing, destroying the original commercial intent and perceptual similarity. To address this, we present SSR-A, a framework tailored for the minimalist rectification of non-compliant image ads. Instead of fine-tuning image editing models directly, SSR-A focuses on translating violation policies into targeted editing instructions. We first introduce a Spatial- and Semantic-Aware Instruction Synthesis Pipeline, where MLLMs synthesize candidate instructions—incorporating spatial grounding and semantic guidance—and select the optimal instruction via multi-dimensional evaluation. Furthermore, we align the model using Curriculum Reinforcement Learning, employing GRPO with multi-faceted rewards to progressively navigate the trade-off between compliance and visual preservation. Extensive experiments and online A/B tests show that SSR-A significantly outperforms state-of-the-art baselines in both compliance and preservation of visual and commercial consistency.

1 Introduction

Online advertising is fundamental to the digital economy, enabling global business growth and sustainability(Rathee and Milfeld, 2024; Campbell et al., 2025). To maintain a healthy and compliant environment, platforms enforce online violation detection systems that screen every ad material(Ji et al., 2025a; Madio and Quinn, 2025; Al Kurdi and Alshurideh, 2025; Ji et al., 2025b). While highly effective at blocking violations, they operate as binary gatekeepers—rejecting millions of non-compliant ads daily. These rejected creatives hold massive untapped commercial value(Ji et al., 2025a). For advertisers, ad rejection creates a costly bottleneck, as manual fixing is labor-intensive, time-consuming, and unscalable(Guo et al., 2021). Consequently, there is a critical need for automated solutions that can bring ads into compliance while preserving their commercial intent.

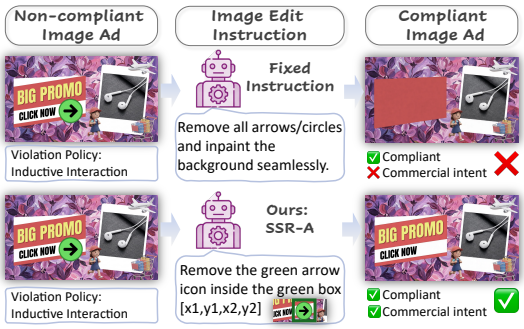


Figure 1: Visualization of the rectification process. The input image (left) contains a non-compliant inductive guidance. **Fixed Instruction** (top) removes the arrow but erases marketing text. **SSR-A** (bottom) precisely removes the arrow while preserving text, demonstrating superior perceptual similarity.

Automated ad rectification can be broadly categorized into textual(Laugier et al., 2021) (e.g., ad copy, OCR text) and visual (e.g., images, video frames) modalities. Although textual rectification has seen rapid progress due to the maturity of Large Language Models (LLMs)(Touvron et al., 2023; OpenAI, 2023), visual rectification remains a significantly more formidable challenge. Existing approaches to responsible image generation generally fall into two paradigms, yet neither is well-suited for the strict constraints of commercial ad rectification. The first paradigm involves iterative generation, exemplified by SafeEditor (Zhang

[*] Equal contribution.
[†] Project leader.
[‡] Corresponding author.

et al., 2025b). These systems iteratively modify the user prompt or latent representation until the output satisfies safety constraints. However, they suffer from an "over-editing" problem: in the pursuit of safety, the model often hallucinates new content or drastically alters the composition, resulting in an image that deviates significantly from the user's original intent. The second paradigm focuses on concept erasure (Sreelatha et al., 2025; Ni et al., 2023; Schramowski et al., 2023), which aims to suppress specific visual concepts (e.g., nudity) during generation. While effective for static definitions, these methods rely on pre-defined lists of forbidden concepts. In the advertising domain, however, violation scenarios are highly context-dependent and evolve rapidly, rendering static concept definitions impractical.

Image ad rectification aims to automatically rectify non-compliant content in ad material detected by online moderation systems, as shown in Fig 1. However, a critical bottleneck is the semantic gap between violation policies and the concrete prompts required by diffusion-based image editing models, which often struggle to accurately interpret such high-level constraints (Li et al., 2025; Fu et al., 2024). Consequently, the core challenge lies in harnessing the superior understanding and localization capabilities of Multimodal Large Language Models (MLLMs) (Liu et al., 2023; OpenAI, 2025) to translate violation policies into targeted editing instructions. Crucially, this transformation process must guarantee compliance without compromising the original commercial intent or perceptual consistency. This constitutes a multi-objective optimization task characterized by inherent conflicts between objectives, making simultaneous optimization non-trivial. Furthermore, the complexity and rapid evolution of violation policies severely constrain manual annotation efforts, resulting in a scarcity of data for SFT that further exacerbates training difficulties. To tackle these limitations, we propose SSR-A, an MLLM-based framework. Our contributions are summarized as follows:

(1) To address data scarcity, we design a Spatial- and Semantic-Aware Instruction Synthesis pipeline. By leveraging MLLMs to generate semantically-guided and spatially-grounded candidates and applying multi-dimensional selection, we curate a high-quality dataset for Supervised Fine-tuning.

(2) To ensure both compliance and visual preservation, we introduce a Curriculum Reinforcement Learning (Bengio et al., 2009; Ko et al., 2022) strategy via Group Relative Policy Optimization (GRPO) (Zhihong Shao, 2024). This approach progressively aligns the model using multi-faceted rewards to achieve compliant, precise editing control.

(3) Integrating these contributions, we present SSR-A, an MLLM-based ad rectification framework. It effectively translates violation policies into targeted editing instructions to enforce compliance while maintaining the original commercial intent. We also release a dataset[1] to facilitate research.

Extensive experiments including online A/B test and ablation studies demonstrate that SSR-A significantly outperforms baselines in terms of both compliance and visual preservation, validating the effectiveness of our proposed framework.

2 Related Work

2.1 MLLM-Guided Image Editing

Recent advancements in image editing (Brooks et al., 2023; Chen et al., 2024; Fu et al., 2023) have shifted from simple text-matching to leveraging the reasoning capabilities of MLLMs. Methods like MGIE (Fu et al., 2024) and Lego-Edit (Jia et al., 2025) utilize MLLMs to interpret user commands and orchestrate editing tools, enabling adaptation to open-domain tasks. To handle ambiguity, EditThinker (Li et al., 2025) introduces a "thinking" process with multi-round reflection, allowing the model to refine its strategy before execution. Similarly, RePlan (Qu et al., 2025) adopts a "plan-then-execute" paradigm, explicitly grounding instructions to specific regions via bounding boxes to mitigate spatial errors. However, these frameworks primarily focus on following explicit user prompts. They struggle with the semantic gap present in ad rectification, where the input is an underspecified violation policy rather than a concrete editing instruction. In this paper, SSR-A introduces a Spatial- and Semantic-Aware Instruction Synthesis strategy, specifically designed to bridge this gap.

2.2 Advertising Content Generation and Compliance

Advertisements demand a delicate balance between aesthetic appeal and strict platform compliance. Recent generative works (Chen et al., 2025; Hu et al., 2025; Deng et al., 2025; Zhang et al., 2025a) primarily focus on synthesizing high-quality content to maximize commercial metrics, typically creating

[1] https://huggingface.co/datasets/rosalie12345/AdRectification_AIGC

assets from scratch rather than preserving existing intent. In parallel, Ji et al. (2025a,b) establish a robust baseline for violation detection and localization but stops short of rectification.

2.3 Safety-Driven Image Rectification

Existing editing paradigms struggle to address this rectification gap effectively. While methods like RePainter (Guo et al., 2025) have made strides in precise e-commerce editing via GRPO, they do not explicitly address compliance constraints. Conversely, safety-driven approaches like SafeEditor (Zhang et al., 2025b) target general risks (e.g., NSFW) but often suffer from "over-editing," compromising the original commercial intent through aggressive modifications. In this paper, we align the model using Curriculum Reinforcement Learning, employing GRPO with multi-faceted rewards to progressively navigate the trade-off between compliance and visual preservation.

3 Method

3.1 Problem Overview

We formulate image ad rectification as a constrained image editing problem. Given a non-compliant image I_{src} and its violation policy V, our goal is to generate a target image I_{tgt} that rectifies V while maximally preserving the commercial intent and perceptual similarity of I_{src}. Considering that diffusion-based models struggle to accurately interpret violation policies, we focus on optimizing an MLLM π_θ to translate violation policies into targeted editing instructions $P \sim \pi_\theta(\cdot \mid I_{src}, V)$. These instructions guide a frozen image editing model $\mathcal{E}$ to generate the compliant output I_{tgt}.

3.2 Spatial- and Semantic-Aware Instruction Synthesis Pipeline (SSIS)

Constructing a high-quality supervised dataset for image ad rectification is challenging due to the complexity of manual annotation and the ambiguity of visual violations (Ji et al., 2025b). To circumvent these limitations, we leverage an advanced MLLM Gemini3-Pro (Google, 2025) as a data engine and propose a *risk-grounded hybrid synthesis framework incorporating multi-dimensional feedback* to autonomously synthesize a robust SFT dataset.

Spatial- and Semantic-Aware Instruction. We observe that standard editing models often yield unsatisfactory results when relying solely on semantic text instructions. For instance, generic commands like "remove the arrows" frequently lead to inaccurate rectification, where models may miss small targets or unintentionally alter the background. While explicit spatial localization can mitigate such issues, rigid spatial constraints are not always optimal for global adjustments—such as altering a character's pose or attire—which rely on holistic semantic understanding. To ensure robust rectification across these scenarios, we introduce a Spatial- and Semantic-Aware Instruction module, as illustrated in Fig 2. This module synthesizes rectification instructions from two synergistic perspectives:

- **Spatially-Grounded ($P_{\mathcal{B}}$):** The instruction explicitly anchors the rectification by incorporating bounding *Box* coordinates (e.g., $[x_1, y_1, x_2, y_2]$) to delineate the violation area.
- **Semantically-Guided ($P_{\mathcal{T}}$):** The instruction relies on purely semantic *Text* descriptions to guide the editor, which is suitable for global or context-dependent modifications.

Multi-dimensional Evaluation and Selection. To ensure data reliability, we generate candidate rectifications via the frozen editor $\mathcal{E}$ and assess them through a collaborative filtering strategy. Specifically, we employ an MLLM to assign a comprehensive quality score based on: (1) *Visual Integrity*, ensuring the absence of visual distortions and the preservation of critical commercial elements (e.g., promotional text); and (2) *Instruction Precision*, verifying the accuracy of box delineation and the logical correctness of the rectification instructions. Evaluation details are provided in the Appendix A. Simultaneously, the online violation detection model assesses (3) *Policy Compliance* to confirm the successful rectification of the violation. Candidates failing the compliance check or falling below an MLLM score threshold are discarded. From the valid set, we select the final instruction P^* by minimizing the LPIPS (Zhang et al., 2018) perceptual distance to the source image. This strategy guarantees that the SFT training data is policy-compliant while maximally preserving the original commercial intent and perceptual similarity.

3.3 Training

3.3.1 Supervised Fine-Tuning

We utilize Qwen3-VL-8B (Bai et al., 2025) as our backbone, and employ SFT to align the model's responses with the specific requirements of violation

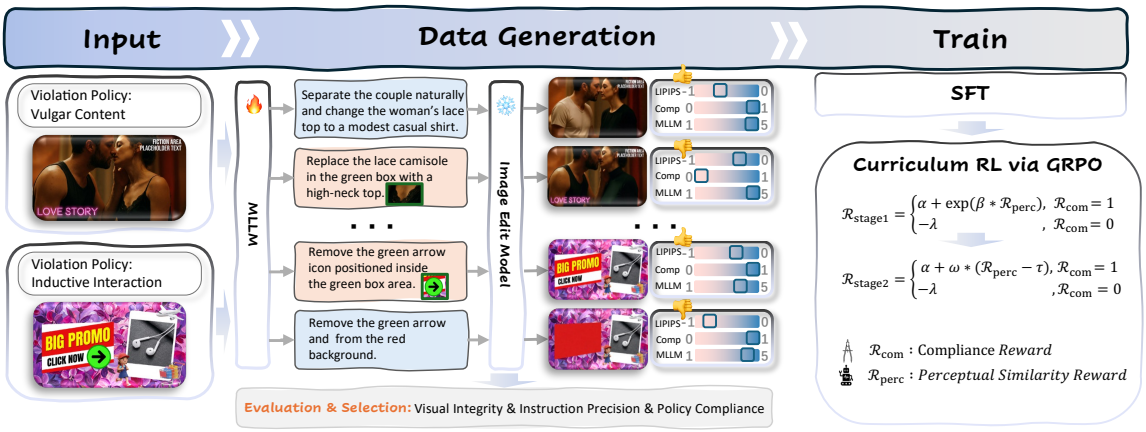


Figure 2: Overview of **SSR-A**. Given inputs (left), the Instruction Synthesis Pipeline (middle) generates spatially-grounded and semantically-guided candidates, filtered via multi-dimensional selection. Finally, the model is optimized through SFT and Curriculum RL (right) to ensure compliance and visual preservation.

rectification. We fine-tune the model on the high-quality paired data $\mathcal{D}_{\text{sft}} = \{(I_{src}, V, P^*)\}$. This stage provides a warm-start for the policy model, ensuring a solid initialization for efficient exploration during Reinforcement Learning (RL).

3.3.2 Reward Functions

We formulate two reward signals to balance the objectives of compliance and the preservation of perceptual similarity:

Compliance Reward (R_{com}). Derived from an online violation detection system, this binary signal serves as a hard constraint for editing validity. We define $R_{\text{com}} \in \{0, 1\}$, where 1 signifies that the image is compliant and 0 indicates a violation.

Perceptual Similarity Reward (R_{perc}). We employ the LPIPS metric to ensure high perceptual similarity between I_{tgt} and I_{src}. The reward is defined as the negative perceptual distance: $R_{\text{perc}} = -\text{LPIPS}(I_{\text{src}}, I_{\text{tgt}})$. Beyond visual consistency, this reward serves as a proxy for commercial intent preservation; our empirical results in Section 4.3 demonstrate that minimizing perceptual deviation effectively preserves the original commercial intent by restricting unnecessary modifications.

3.3.3 Curriculum Reinforcement Learning (CRL)

While SFT provides strong initialization, it relies on static supervision, which struggles to balance the conflicting objectives of compliance and visual preservation. Therefore, we employ a two-stage curriculum RL via GRPO with a hierarchical reward system, as illustrated in Fig 2.

Stage 1: Compliance-Prioritized Exploration (CPE). The primary objective of the initial stage is to guide policy model to learn that satisfying the compliance constraint ($R_{\text{com}} = 1$) is the absolute prerequisite. The reward function is defined as:

$$R_{\text{stage1}} = \begin{cases} \alpha + \exp(\beta \cdot R_{\text{perc}}), & \text{if } R_{\text{com}} = 1 \\ -\lambda, & \text{if } R_{\text{com}} = 0 \end{cases} \tag{1}$$

We utilize the exponential function for its gradient property: $\nabla \exp(x) \to 0$ as $x \to -\infty$ (i.e., high LPIPS). For hard cases requiring significant edits, R_{perc} is inherently low. A linear penalty would generate strong negative gradients, potentially discouraging the model from making necessary edits to achieve compliance. By using an exponential term, the marginal penalty diminishes rapidly for large edits, effectively providing a "soft constraint". This tolerance encourages the model to prioritize compliance without being overly penalized for the inevitable visual changes in the early training phase. The base reward of α ensures that any compliant generation is strictly preferred over a violation.

Stage 2: Perception-Preserving Refinement (PPR). Once compliance stabilizes, we address the "vanishing gradient" problem of the exponential function. In Stage 1, the model may "give up" on optimizing hard cases because the gradient signal vanishes when LPIPS is high. To force continuous refinement, we switch to a piecewise linear reward:

$$R_{\text{stage2}} = \begin{cases} \alpha + \omega \cdot (R_{\text{perc}} - \tau), & \text{if } R_{\text{com}} = 1 \\ -\lambda, & \text{if } R_{\text{com}} = 0 \end{cases} \tag{2}$$

Method	Industrial Dataset										Open-Source Dataset									
	Inductive Interaction			Sensitive Content			Average				Inductive Interaction			Sensitive Content			Average			
	CR(%)↑	VQ↑	LPIPS↓	CR(%)↑	VQ↑	LPIPS↓	CR(%)↑	VQ↑	LPIPS↓	Overall↑	CR(%)↑	VQ↑	LPIPS↓	CR(%)↑	VQ↑	LPIPS↓	CR(%)↑	VQ↑	LPIPS↓	Overall↑
Part I: Evaluation of Instruction Generation Strategies																				
Gemini3-Pro with $P_{\mathcal{T}}$	99.0	4.39	0.210	83.4	4.71	0.183	86.9	4.64	0.189	78.1	99.5	4.62	0.177	83.6	4.86	0.158	87.3	4.81	0.163	82.1
Gemini3-Pro with $P_{\mathcal{B}}$	97.0	4.45	0.187	86.5	4.64	0.193	88.9	4.59	0.192	78.3	98.9	4.53	0.196	85.1	4.81	0.173	88.3	4.74	0.178	80.9
Gemini3-Pro with SSIS	99.5	4.38	0.154	91.0	4.68	0.172	92.9	4.61	0.168	**87.2**	99.5	4.66	0.152	88.2	4.87	0.150	90.8	4.82	0.150	**87.7**
Part II: Comparison with Baselines and Training Stages																				
Fixed Instruction	98.0	4.40	0.257	86.2	4.79	0.287	88.9	4.70	0.281	57.5	96.2	4.53	0.202	86.7	4.89	0.254	88.9	4.81	0.242	64.5
Qwen3-VL-8B	97.5	4.51	0.226	85.8	4.71	0.216	88.4	4.67	0.218	73.1	97.3	4.65	0.188	88.2	4.86	0.191	90.3	4.81	0.191	79.9
Ours (SFT)	98.5	4.50	0.175	79.2	4.66	0.196	83.6	4.62	0.191	74.5	99.5	4.55	0.179	79.1	4.84	0.176	83.8	4.77	0.177	77.4
Ours (SSR-A)	98.5	4.41	0.159	84.9	4.64	0.184	88.0	4.59	0.178	**79.5**	99.5	4.55	0.165	83.0	4.82	0.162	86.8	4.76	0.162	**81.6**

Table 1: Quantitative results on Industrial and Open-Source Datasets.

where α and λ are stage-specific hyperparameters (detailed in Appendix B), and $\tau = -0.22$ is a tolerance threshold for perceptual change. Unlike the exponential curve, this linear form provides a constant gradient flow, ensuring the model continues to optimize even when the consistency score is low. The mechanism adapts to two scenarios: For **Minor Refinements** ($R_{\text{perc}} \geq \tau$), the term $(R_{\text{perc}} - \tau)$ is positive. We apply a strong weight $\omega = 5.0$ to force the model to retain original details as much as possible when only slight edits are needed. For **Major Alterations** ($R_{\text{perc}} < \tau$), the term becomes negative. Here, we reduce the weight to $\omega = 2.0$. This lenient penalty ensures the final reward remains positive even with large visual changes, preventing the model from abandoning the compliance constraint due to excessive consistency punishment.

4 Experiments

4.1 Experimental Setup

Evaluation Datasets. We utilize two datasets for evaluation: (1) An internal real-world ad-compliance dataset consisting of 1k non-compliant advertising images, categorized into Inductive Interaction (e.g., fake directional arrows), Tattoos, Tobacco & Alcohol, and Vulgar Content. (2) An open-source evaluation dataset constructed via Image-to-Text-to-Image. This dataset desensitizes the industrial data for reproducibility while preserving original violation patterns. Detailed distributions are provided in the Appendix C.

Baselines. We evaluate our framework in two parts. First, to assess data quality, we compare instruction generation strategies on Gemini3-Pro: using semantically-guided prompts $P_{\mathcal{T}}$, spatially-grounded prompts $P_{\mathcal{B}}$, and our proposed SSIS. Second, to validate our training stages, we compare our method against: (1) Fixed Instruction, a static prompt baseline; (2) Qwen3-VL-8B, a strong zero-shot MLLM; and (3) our training stages, including Ours (SFT) and Ours (SSR-A).

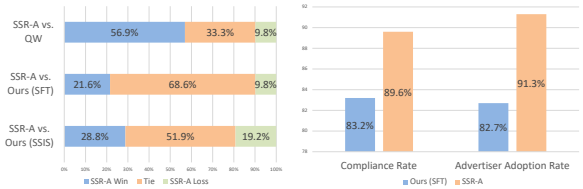
Metrics. To comprehensively evaluate the rectification performance, we employ four key metrics: (1) **Compliance Rate (CR):** The percentage of samples which are compliant; (2) **LPIPS:** A metric reflecting the preservation of the original visual content; (3) **Visual Quality (VQ):** We employ Qwen3-VL-30B to detect deformations or artifacts. This metric serves as a critical monitor for "reward hacking", ensuring that the model does not achieve compliance through trivial solutions—such as destroying image structure or removing essential subjects; (4) **Overall:** A composite score designed to provide a holistic assessment. It integrates CR and LPIPS to identify the optimal operating point approaching the Pareto frontier between compliance and visual preservation. Detailed calculation protocols are provided in Appendix D.

4.2 Main Results

Effectiveness of SSIS. Part I of Table 1 validates the superiority of our hybrid synthesis pipeline. Compared to strategies relying solely on spatial ($P_{\mathcal{B}}$) or semantic ($P_{\mathcal{T}}$) guidance, the SSIS approach achieves the optimal balance between compliance and visual consistency. Specifically, it attains a leading Overall score of **87.2** on the industrial dataset. This confirms that leveraging multi-dimensional evaluation and selection is essential for synthesizing high-quality training data.

Effectiveness of SSR-A. Part II of Table 1 demonstrates the efficacy of our SSR-A. While baselines like Fixed Instruction and Qwen3-VL-8B achieve high CR, they suffer from severe over-editing (high LPIPS). In contrast, SSR-A significantly refines the SFT baseline, improving the Overall score from 74.5 to **79.5** on the Industrial Dataset. This demonstrates that SSR-A effectively navigates the Pareto frontier, achieving precise rectification without compromising visual consistency.

Note that all experiments in this section utilize



(a) Human Preference (b) Online Testing

Figure 3: **Quantitative results of user study and online A/B test. (a)** Our method achieves a 2.2× win ratio against the SFT baseline in the human preference study. **(b)** Online A/B testing demonstrates significant gains in both compliance and utility.

a proprietary, ad-optimized diffusion model as the backbone editor. To verify generalizability, additional validations on open-source backbones are provided in Appendix E. Additionally, a qualitative analysis of rectification outcomes—including per-category difficulty, typical failure modes, and LPIPS score distributions—is provided in Appendix F.

4.3 User Study and Online A/B Testing

To evaluate real-world applicability, we conduct both a User Study and an Online A/B Test.

4.3.1 User Study

While quantitative metrics like LPIPS measure perceptual similarity, they cannot fully assess the preservation of the original commercial intent. In order to bridge this gap, we conducted a rigorous blind pairwise study with five professional advertisers on 100 randomly sampled compliant rectifications. We specifically compared SSR-A against Qwen3-VL-8B, Ours (SFT), and Ours (SSIS).

As shown in Fig 3a, SSR-A consistently achieves the highest preference rates. Notably, the high tie rate (68.6%) against SFT suggests that while SFT handles routine cases adequately, SSR-A dominates in the remaining non-tie scenarios. This confirms that our reinforcement strategy effectively guarantees strict compliance while preserving the source ad's commercial intent.

4.3.2 Online A/B Test

To validate real-world impact, we deployed SSR-A on a large-scale advertising platform for a 3-day A/B test, allocating 20% of live traffic against the Ours (SFT) baseline. We evaluated performance using *Advertiser Adoption Rate* (proxy for commercial intent) and *Compliance Rate*. In Fig 3b, SSR-A yielded a substantial **+8.6%** increase in adoption and **+6.4%** in compliance over SFT. These results confirm that our framework successfully balances strict compliance requirements with commercial value in a production environment.

Method	CR (%) ↑	VQ ↑	LPIPS ↓	Overall ↑
SFT	83.56	4.62	0.1911	74.5
RL with PPR	84.33	4.56	**0.1753**	76.3
RL with CPE	86.89	**4.63**	0.1846	77.8
RL with CRL	**88.00**	4.59	0.1781	**79.5**

Table 2: Ablation study on the curriculum RL stages.

4.4 Ablation Study on Curriculum Strategy

To validate the necessity of our two-stage reward design, we conducted an ablation study comparing the SFT baseline against variants using single-stage RL strategies. The results in Table 2 demonstrate that our CRL effectively balances the conflicting objectives of compliance and visual preservation.

Impact of Linear Reward (PPR Only). PPR yields the lowest LPIPS (0.1753) but only a marginal CR improvement over SFT (84.33% vs. 83.56%). This confirms that the linear reward imposes strict penalties on visual changes. For samples requiring significant edits, strong negative gradients discourage necessary structural alterations, trapping the model in a local optimum prioritizing visual consistency over compliance.

Impact of Exponential Reward (CPE Only). CPE boosts CR to 86.89% but yields higher LPIPS (0.1846). Acting as a "soft constraint," the exponential reward encourages large edits via diminishing penalties. However, vanishing gradients ($\nabla \exp(x) \to 0$) fail to provide sufficient signal for fine-tuning visual details once compliance is met, limiting perceptual similarity preservation.

Effectiveness of Curriculum Learning (CRL). CRL achieves the optimal trade-off, securing the highest CR (**88.00%**) and Overall Score (**79.5**). By initially employing CPE (Stage 1), we establish a robust compliance foundation. Subsequently, switching to PPR (Stage 2) reintroduces constant gradient flow to recover visual details. This progressive optimization guarantees strict compliance while maximizing the preservation of the original perceptual similarity.

5 Limitations

Our framework has several limitations. First, the SSR-A pipeline involves sequential MLLM

inference, diffusion-based editing, and multi-dimensional evaluation, which incurs non-trivial computational overhead. This makes it better suited for asynchronous offline repair of rejected advertisements than for real-time, latency-sensitive editing scenarios. Second, while our user study demonstrates overall preference for SSR-A, the current framework does not account for individual advertiser-specific aesthetic requirements, which we leave for future work. Third, the current framework is designed exclusively for static image ads; extending it to video advertisements or multi-frame content, which requires temporal consistency and inter-frame coherence, remains future work.

6 Conclusion

We presented SSR-A, an MLLM-based framework for rectifying non-compliant ad images. Addressing over-editing issues, SSR-A balances policy compliance with commercial intent preservation. This is achieved by bridging the semantic gap via a Spatial- and Semantic-Aware Instruction Synthesis pipeline and optimizing alignment through Curriculum Reinforcement Learning with GRPO. Experiments and online A/B tests confirm SSR-A significantly outperforms state-of-the-art baselines.

7 Ethical Considerations

This work is conducted in strict accordance with established ethical guidelines and data privacy regulations. The evaluation dataset we release consists exclusively of AI-synthesized images produced by diffusion models from textual descriptions, and contains no authentic advertisements or personally identifiable information. Any non-compliant examples presented in this paper are included solely for illustrative purposes to facilitate scientific analysis. All released resources are intended exclusively for non-commercial academic research.

References

Barween Al Kurdi and Muhammad Turki Alshurideh. 2025. The effect of social media influencer traits on consumer purchasing decisions for keto products: examining the moderating influence of advertising repetition. *Journal of Marketing Communications*, 31(4):422–443.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, and 45 others. 2025. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM.

Tim Brooks, Aleksander Holynski, and Alexei A Efros. 2023. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18392–18402.

Colin Campbell, Sean Sands, Brent McFerran, and Alexis Mavrommatis. 2025. Diversity representation in advertising. *Journal of the Academy of Marketing Science*, 53(2):588–616.

SiXiang Chen, Jianyu Lai, Jialin Gao, Tian Ye, Haoyu Chen, Hengyu Shi, Shitong Shao, Yunlong Lin, Song Fei, Zhaohu Xing, and 1 others. 2025. Postercraft: Rethinking high-quality aesthetic poster generation in a unified framework. *arXiv preprint arXiv:2506.10741*.

Xi Chen, Lianghua Huang, Yu Liu, Yujun Shen, Deli Zhao, and Hengshuang Zhao. 2024. Anydoor: Zero-shot object-level image customization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6593–6602.

Jiaxin Deng, Shiyao Wang, Kuo Cai, Lejian Ren, Qigen Hu, Weifeng Ding, Qiang Luo, and Guorui Zhou. 2025. Onerec: Unifying retrieve and rank with generative recommender and iterative preference alignment. *arXiv preprint arXiv:2502.18965*.

Tsu-Jui Fu, Wenze Hu, Xianzhi Du, William Yang Wang, Yinfei Yang, and Zhe Gan. 2023. Guiding instruction-based image editing via multimodal large language models. *arXiv preprint arXiv:2309.17102*.

Tsu-Jui Fu, Wenze Hu, Xianzhi Du, William Yang Wang, Yinfei Yang, and Zhe Gan. 2024. Guiding Instruction-based Image Editing via Multimodal Large Language Models. In *International Conference on Learning Representations (ICLR)*.

Google. 2025. A new era of intelligence with gemini 3. https://blog.google/products/gemini/gemini-3.

Shunan Guo, Zhuochen Jin, Fuling Sun, Jingwen Li, Zhaorui Li, Yang Shi, and Nan Cao. 2021. Vinci: An intelligent graphic design system for generating advertising posters. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

Zipeng Guo, Lichen Ma, Xiaolong Fu, Gaojing Zhou, Lan Yang, Yuchen Zhou, Linkai Liu, Yu He, Ximan Liu, Shiping Dong, and 1 others. 2025. Repainter: Empowering e-commerce object removal via spatial-matting reinforcement learning. *arXiv preprint arXiv:2510.07721*.

Xiwei Hu, Haokun Chen, Zhongqi Qi, Hui Zhang, Dexiang Hong, Jie Shao, and Xinglong Wu. 2025. Dreamposter: A unified framework for image-conditioned generative poster design. *arXiv preprint arXiv:2507.04218*.

Deyi Ji, Yuekui Yang, Liqun Liu, Peng Shu, Haiyang Wu, Shaogang Tang, Xudong Chen, Shaoping Ma, Tianrun Chen, and Lanyun Zhu. 2025a. RAVEN++: Pinpointing fine-grained violations in advertisement videos with active reinforcement reasoning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1–10.

Deyi Ji, Yuekui Yang, Haiyang Wu, Shaoping Ma, Tianrun Chen, and Lanyun Zhu. 2025b. RAVEN: Robust advertisement video violation temporal grounding via reinforcement reasoning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pages 22–31.

Qifei Jia, Yu Liu, Yajie Chai, Xintong Yao, Qiming Lu, Yasen Zhang, Runyu Shi, Ying Huang, and Guoquan Zhang. 2025. Lego-edit: A general image editing framework with model-level bricks and mllm builder. *arXiv preprint arXiv:2509.12883*.

Keonwoo Ko, Gu Jin, and Youngchul Sung. 2022. Curriculum offline reinforcement learning. In *International Conference on Learning Representations*.

Léo Laugier, John Pavlopoulos, Jeffrey Sorensen, and Lucas Dixon. 2021. Civil rephrases of toxic texts with self-supervised transformers. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1442–1461.

Hongyu Li, Manyuan Zhang, Dian Zheng, Ziyu Guo, Yimeng Jia, Kaituo Feng, Hao Yu, Yexin Liu, Yan Feng, Peng Pei, and 1 others. 2025. Editthinker: Unlocking iterative reasoning for any image editor. *arXiv preprint arXiv:2512.05965*.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning.

Leonardo Madio and Martin Quinn. 2025. Content moderation and advertising in social media platforms. *Journal of Economics & Management Strategy*, 34(2):342–369.

Minheng Ni, Chenfei Wu, Xiaodong Wang, Shengming Yin, Lijuan Wang, Zicheng Liu, and Nan Duan. 2023. Ores: Open-vocabulary responsible visual synthesis. *arXiv preprint arXiv:2308.13785*.

OpenAI. 2023. Gpt-4 technical report. Technical report, OpenAI.

OpenAI. 2025. GPT-5 technical report. Forthcoming.

Tianyuan Qu, Lei Ke, Xiaohang Zhan, Longxiang Tang, Yuqi Liu, Bohao Peng, Bei Yu, Dong Yu, and Jiaya Jia. 2025. Replan: Reasoning-guided region planning for complex instruction-based image editing. *arXiv preprint arXiv:2512.16864*.

Shelly Rathee and Tyler Milfeld. 2024. Sustainability advertising: literature review and framework for future research. *International Journal of Advertising*, 43(1):7–35.

Patrick Schramowski, Manuel Brack, Björn Deiseroth, and Kristian Kersting. 2023. Safe latent diffusion: Mitigating inappropriate degeneration in diffusion models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Silpa Vadakkeeveetil Sreelatha, Sauradip Nag, Muhammad Awais, Serge Belongie, and Anjan Dutta. 2025. Respodiff: Dual-module bottleneck transformation for responsible faithful t2i generation. *Preprint*, arXiv:2509.15257.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhengyi Wang, An Yang, Bowen Yu, Chen Cheng, Dayiheng Liu, Deqing Li, and 20 others. 2025. Qwen-image technical report. *Preprint*, arXiv:2508.02324.

Jun Zhang, Yi Li, Yue Liu, Changping Wang, Yuan Wang, Yuling Xiong, Xun Liu, Haiyang Wu, Qian Li, Enming Zhang, and 1 others. 2025a. Gpr: Towards a generative pre-trained one-model paradigm for large-scale advertising recommendation. *arXiv preprint arXiv:2511.10138*.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*.

Ruiyang Zhang, Jiahao Luo, Xiaoru Feng, Qiufan Pang, Yaodong Yang, and Juntao Dai. 2025b. Safeeditor: Unified mllm for efficient post-hoc t2i safety editing. *arXiv preprint arXiv:2510.24820*.

Qihao Zhu Runxin Xu Junxiao Song Mingchuan Zhang Y.K. Li Y. Wu Daya Guo Zhihong Shao, Peiyi Wang. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models.

A Details of Multi-Dimensional Evaluation

Please refer to Fig 5 for the detailed implementation of our multi-dimensional evaluation strategy.

The figure showcases the exact system prompt utilized to guide the MLLM in assessing candidate image-instruction pairs. To ensure the quality of the training dataset, we enforce a strict filtering criterion: only samples achieving a comprehensive quality score greater than 4 are retained.

B Implementation Details.

Our method is built upon Qwen3-VL-8B-Instruct. For the reward functions defined in Eq. (1) and Eq. (2), we set the following hyperparameters: In Stage 1 (CPE), $\alpha = 0.5$, $\beta = 3.5$, and $\lambda = 1.0$. In Stage 2 (PPR), $\alpha = 2.0$, $\lambda = 2.0$.

We perform LoRA-based SFT (rank 8, α 32) on our constructed dataset ($\sim$6k samples) with a learning rate of 1×10^{-4} and a global batch size of 64, selecting the optimal checkpoint at 500 steps. Subsequently, we conduct Curriculum GRPO with a learning rate of 1×10^{-5}, a group size (G) of 4, and a KL coefficient of 0.01. All experiments are conducted on a cluster of 4 NVIDIA H20 GPUs.

C Details of the Open-Source Evaluation Dataset

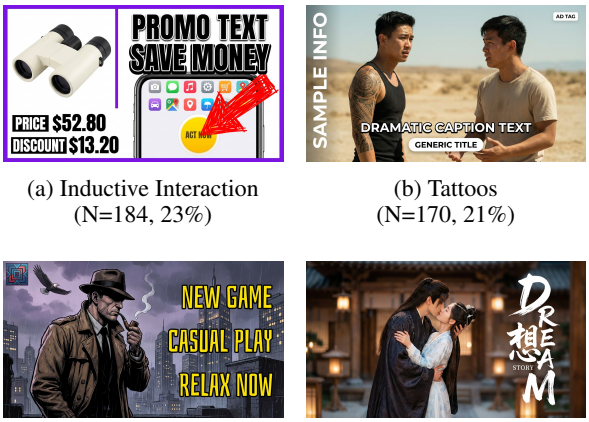
To ensure our dataset faithfully reflects the complexity and distribution of real-world online advertising while strictly adhering to privacy and copyright regulations, we employ a Generative Resynthesis strategy based on precise semantic descriptions. We first extract detailed textual descriptions from real-world online advertising data that contain specific violations. These descriptions are then utilized as prompts for advanced text-to-image diffusion models to synthesize entirely new images. This approach allows us to reproduce the authentic distribution of risk scenarios found in commercial ads, while ensuring that all visual content is synthetically generated and fully anonymized. Sample visualization and detailed statistics of the resulting dataset are presented in Fig 4.

D Metric Details

In this section, we provide the implementation details for our MLLM-based evaluators and the definition of our composite score $Overall$.

D.1 MLLM-based Visual Quality Evaluation

While traditional reference-based metrics like LPIPS quantify pixel-level fidelity, they often fail to capture high-level perceptual artifacts that degrade user experience. To address this, we employ



(a) Inductive Interaction (N=184, 23%)

(b) Tattoos (N=170, 21%)

(c) Tobacco & Alcohol (N=185, 23%)

(d) Vulgar Content (N=261, 33%)

Figure 4: **Sample visualization and distribution.** Displayed in a 2×2 grid for better visibility.

Score	Visual Quality (VQ)
5	**Commercial Grade.** Flawless lighting, texture, and edge consistency. No perceptible editing traces.
4	**High Quality.** Clean removal/repair. Minor, imperceptible noise that does not affect usability.
3	**Acceptable.** Slight blur or minor texture mismatch, but the object is recognizable.
2	**Poor.** Visible artifacts, smearing, or obvious unnatural transitions.
1	**Failure.** Severe distortion, black blocks, or global collapse.

Table 3: **Automated Evaluation Rubric.** We utilize Qwen3-VL-30B to assign scores from 1 to 5 based on these specific visual criteria.

Qwen3-VL-30B as an automated evaluator to assess **Visual Quality (VQ)**.

This metric evaluates the naturalness and coherence of the rectified image, specifically detecting artifacts, blurring, unnatural transitions, or structural collapse. Crucially, VQ serves as a necessary guardrail against "reward hacking," ensuring that the model does not achieve a high compliance rate (CR) by simply deleting the subject or destroying the image structure. The detailed scoring rubric used by the MLLM is presented in Table 3.

D.2 Definition of the Overall Score

To quantitatively evaluate the method's performance in balancing compliance with visual preservation, we utilize a composite score $Overall$. For each test sample, the score is defined as:

$$S = \mathbb{I}_{\text{com}} \cdot \frac{100}{1 + \exp\left(k \cdot (\mathcal{L} - \tau)\right)} \tag{3}$$

where $\mathbb{I}_{\text{com}}$ is a binary indicator that equals 1

if the image satisfies safety compliance standards and 0 otherwise. This hard constraint reflects the practical reality that non-compliant advertisements are unusable for commercial applications. $\mathcal{L}$ denotes the LPIPS. The logistic term maps the LPIPS value to a normalized score range of $[0, 100]$, where $\tau = 0.3$ and $k = 150$ regulate the midpoint and steepness of the penalty curve, respectively.

E Generalizability of SSR-A

To assess the generalizability of our framework, we directly applied our trained MLLM to guide an unseen open-source editing model, **Qwen-Image-Edit** (Wu et al., 2025), without any additional training. We compare this zero-shot transfer setting against a baseline where Gemini3-Pro generates instructions for the same editor. As shown in Table 4, our method outperforms the Gemini3-Pro baseline across all metrics: achieving higher safety compliance (72.67% vs. 72.11%), significantly better visual consistency (0.1509 vs. 0.1816 LPIPS), and a superior overall score (66.6 vs. 64.9). This demonstrates that our framework learns robust, model-agnostic strategies that can effectively control diverse downstream editors.

Method (MLLM + Editor)	CR (%) ↑	LPIPS ↓	Overall ↑
Gemini3-Pro + Qwen-Image-Edit	72.11	0.1816	64.9
SSR-A + Qwen-Image-Edit	**72.67**	**0.1509**	**66.6**

Table 4: Zero-shot Transfer to Qwen-Image-Edit.

F Qualitative Analysis of Rectification Outcomes

To provide deeper insight into the behavior of SSR-A beyond aggregate metrics, we analyze rectification outcomes across violation categories. Inductive Interaction is relatively straightforward, as the violating elements (e.g., fake directional arrows) are typically well-localized and visually separable from commercial content, allowing precise removal with minimal collateral changes. In contrast, Sensitive Content categories—including Tattoos, Tobacco & Alcohol, and Vulgar Content—are considerably more challenging, often requiring semantically complex edits such as modifying human poses, replacing attire, or substituting objects, which span larger spatial extents and demand higher-level scene understanding. At a finer granularity, the difficulty ranking across subcategories is: *Vulgar Content* > *Inductive Interaction* > *Tobacco & Alcohol* > *Tattoos*. In the majority of cases, SSR-A successfully resolves violations while preserving commercial intent, benefiting from clear spatial separation between violating and commercial regions. The most common difficulty arises when the two are spatially entangled—for instance, removing an inductive arrow may inadvertently delete overlapping marketing text, and rectifying vulgar content in human figures sometimes alters adjacent product displays or brand logos. Examining the distribution of LPIPS scores further reveals three characteristic clusters: non-compliant outputs cluster below 0.10, indicating insufficient editing that preserved visual similarity at the expense of compliance; the majority of successful rectifications fall within $0.10 - 0.35$, reflecting an effective trade-off; and a small number of outputs exceed 0.50, where compliance was achieved at the cost of substantial visual degradation.

G Additional Qualitative Results

To provide a more comprehensive evaluation of our proposed framework, we present additional qualitative comparisons in Fig 6. These extended case studies cover a broader range of safety violation policies and diverse visual scenarios. We compare our method against the baselines (*Fixed Instruction*, *Qwen3-VL-8B*) and ablation variants (*SFT*, *SSIS*) detailed in the main text.

Role: You are an Ad Image Restoration Quality Assurance Expert with 20 years of experience. Your task is to execute a "Multi-Faceted Evaluation" to assess the quality of AI-repaired advertising images based on strict academic definitions.

Task: Compare the Image Before and Image After, considering the Violation Reason and Original Instruction. Evaluate the result based on two core dimensions: **Visual Integrity** and **Instruction Precision**, and assign a single **Comprehensive Quality Score (1-5)**.

Evaluation Dimensions

Visual Integrity

Definition: Ensuring artifact-free generation and the retention of critical commercial elements.

Core Focus: Is the image clean, high-fidelity, and commercially safe?

Key Checks:

Are there artifacts like "Green Box Residue", blurring, or distortion?

Are non-violation areas (advertising text, brand logos, product details) perfectly preserved?

Instruction Precision

Definition: Verifying the precision of box delineation and the logical coherence of the instructions.

Core Focus: Was the editing area precise? Is the result logically sound?

Key Checks:

Is the modification strictly confined to the violation area (not too large/small)?

Does the generated content fit the narrative logic, lighting, and physics of the original scene?

Scoring Criteria (1-5)

5 (Commercial Grade): Flawless. The image is artifact-free. All commercial elements (text/logos) are pixel-perfect. The edit is precise and logically coherent.

4 (High Quality): Excellent. Minor, imperceptible noise that does not affect commercial usability. Text and logos are intact.

3 (Acceptable): Usable but imperfect. Slight blur or minor texture mismatch, but the object is recognizable and text is readable.

2 (Poor): Visible artifacts, smearing, or obvious unnatural transitions. Commercial elements (text/logos) may be slightly damaged.

1 (Failure): Severe distortion, black blocks, global collapse, or critical loss of commercial elements (e.g., unreadable text).

Input Data

Violation Reason: {{violation_policy}}

Original Instruction: {{Instruction}}

Images: [Image_Before], [Image_After]

Output Format (JSON)

Please output strictly in JSON format without Markdown tags.

{ "score": 1-5, // Comprehensive Quality Score based on Visual Integrity and Instruction Precision "reasoning": "Brief analysis (under 100 words). Explain the score by referencing both Visual Integrity (e.g., artifacts, text preservation) and Instruction Precision (e.g., box accuracy, logic)." }

Examples

Case 1: Damaged Text (Visual Integrity Failure)

Input: Violation "Deep V". Image: Cleavage covered, but "50% OFF" text is blurred.

Output: { "score": 2, "reasoning": "Although the violation was removed, Visual Integrity is compromised as the '50% OFF' text is blurred, destroying commercial intent. The edit is precise but the result is not commercially usable." }

Case 2: Perfect Rectification

Input: Violation "Tattoo". Image: Tattoo removed, skin texture natural, background logo sharp.

Output: { "score": 5, "reasoning": "Visual Integrity is perfect with no artifacts and sharp background logos. Instruction Precision is high as the edit is strictly confined to the tattoo area with natural skin texture generation." }

Figure 5: System Prompt: Multi-Dimensional Evaluation for Ad Rectification

Figure 6: **Qualitative comparison of visual safety rectification across different methods.** Each row represents a distinct violation policy (e.g., *Tobacco & Alcohol*, *Vulgar Content*). The **first column** shows the input image with the detected violation policy. Subsequent columns display the rectification results from *Fixed Instruction*, *Qwen3-VL-8B*, *Ours (SFT)*, *Ours (SSIS)*, and *Ours (SSR-A)*. Results failing the online safety audit are marked with a **red "Non-compliant"** label below the image. For compliant results, LPIPS scores are provided to quantify perceptual similarity to the original image (lower is better), with the best score in each row **highlighted in green**. The results demonstrate that our method achieves minimalist rectification, resolving violations while maintaining the highest perceptual similarity and visual preservation.