Social-R1: Incentivizing Social Relation Reasoning Capability of Multimodal Large Language Models via Reinforcement Learning

Anonymous EMNLP submission

Abstract

Social relationship recognition, which infers relationship types between individuals, is crucial for the deep understanding of semantically rich multimodal scenarios, supporting a wide range of downstream applications. However, despite advances in classification accuracy achieved by end-to-end learning frameworks and knowledge-enhanced models, current approaches still face challenges in generalization, interpretability, and efficiency. In this paper, we introduce Social-R1, a multimodal large language model trained with reinforcement learning (RL) for social relationship recognition. Our approach enables end-to-end reasoning directly from images and bounding boxes, without requiring multi-stage pipelines or handcrafted prompts. Social-R1 achieves state-of-the-art performance on the PIPA and PISC benchmarks, while generating human-understandable rationales that significantly improve interpretability in social relationship recognition. The code is available at `https://anonymous.4open.science/r/Test-57B7`.

1 Introduction

Social relationships play a pivotal role in our lives, profoundly influencing our emotional, psychological, and physical well-being while forming the foundation for human social activities (Kitayama and Markus, 2000). In this case, social relationship recognition aims to infer the types of social relationships between individuals (e.g., parent-child, romantic partners, colleagues), which provides valuable insights for understanding human daily life. Besides, social relationship information can also facilitate progress in related fields, such as plot analysis, video question answering, and intelligent content distribution.

However, social relationship recognition is a high-level cognitive task that requires integrating rich prior knowledge and semantic information about social relationships (Wang et al., 2018; Li et al., 2020; Guo et al., 2023; Yu et al., 2024). For instance, GRM (Wang et al., 2018) integrates a knowledge graph into its model to leverage contextual object information, while TRGAT (Guo et al., 2023) utilizes logical constraints among multiple social relationships within the same scene. Unfortunately, although these models have demonstrated impressive performance, they continue to be constrained by challenges in generalization and interpretability, i.e., there is no assurance that they will perform reliably in unseen scenarios, nor can they consistently offer transparent or understandable reasoning behind their outputs.

Moreover, with the development of Large Language Models (LLMs), LLM-based solutions have increasingly been adopted for social relationship reasoning. For instance, SocialGPT (Li et al., 2024) constructs "social stories" as intermediate representations and then leverages prompt engineering to reason about social relationships. Although this method does not require fine-tuning on downstream tasks and demonstrates interpretability, its complex multi-stage process and reliance on prompt designs lead to high inference latency and sensitivity to prompt. Recently, Multimodal Large Language Models (MLLMs), which are built upon LLMs and inherit their reasoning potential while possessing the ability to process multimodal inputs, have been used to handle vision classification tasks (Wu et al., 2023; Zhang et al., 2024). Besides, studies (Li et al., 2025; Zhou et al., 2025; Liu et al., 2025a; Zhan et al., 2025) have primarily explored how reinforcement learning (RL) can enhance the performance of MLLMs in specific tasks, such as perception tasks and logical reasoning tasks. However, social relationship reasoning, which is a task that demands both accurate visual perception and detailed reasoning in complex social contexts, remains underexplored.

To address the above challenges, we introduce Social-R1, a novel multimodal LLM with explicit

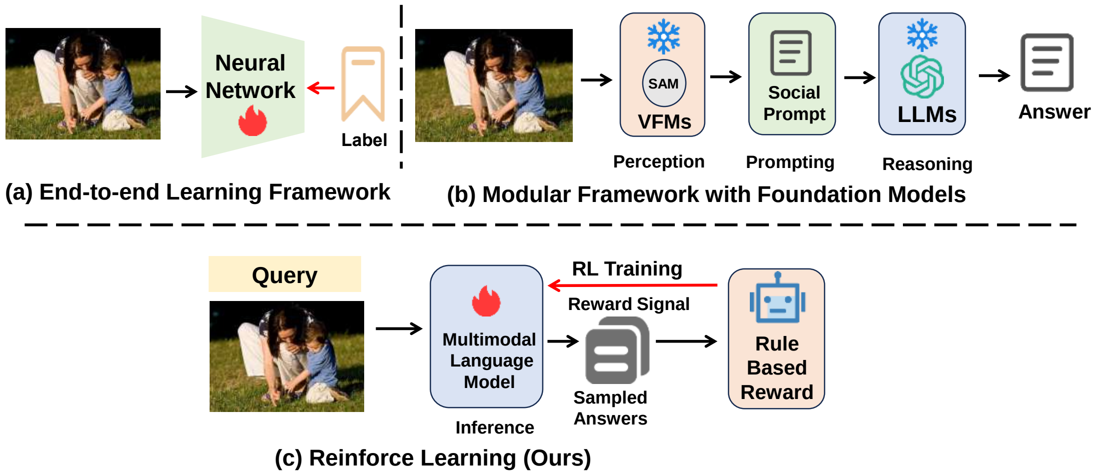


Figure 1: Comparisons of different technical frameworks for social relationship recognition task.

reasoning capabilities that excels at social relationship reasoning tasks. The approach integrates perception and reasoning into a single forward pass by directly processing images with bounding boxes, thus eliminating the need for complex multi-stage pipelines or handcrafted prompts. As shown in Figure 1, we first design a multimodal language model that takes an image and a corresponding query as input and generates multiple sampled answers. These answers are then evaluated by a rule-based reward system, which provides feedback signals to improve model performance. Subsequently, we deploy an RL-based training module that not only optimizes the model for better performance but also enables it to generate human-understandable rationales, thereby enhancing interpretability.

The technical contributions of this paper could be summarized as follows:

- We propose Social-R1, a novel multimodal large language model that excels at social relation reasoning tasks with explicit reasoning capabilities.
- We achieve accurate and interpretable results by employing RL-based methodology and utilizing a rule-based reward system to obtain feedback signals.
- Experiments demonstrate that our method yields the state-of-the-art performance and explainable results on the PIPA and PISC benchmarks.

2 Related Works

2.1 Social Relation Recognition

In recent years, social relationship recognition has received widespread attention. For instance, Sun et al. (2017) followed Bugental's domain-based theory (Bugental, 2000) and annotated the PIPA dataset, which has become one of the most popular benchmarks for social relation recognition. Similarly, Li et al. (2017) adopted the relational models theory (Fiske, 1992) and contributed the People in Social Context (PISC) dataset.

Building upon these datasets, Dual-Glance (Li et al., 2017) introduced a dual glance model, where the first glance focuses on analyzing the pair of individuals of interest, and the second glance captures contextual information derived from objects detected in the scene. Noticng that there are often multiple social relations within the same image, GR^2N (Li et al., 2020) proposed to jointly infer all relations on an image with graph neural networks. Additionally, MGR (Zhang et al., 2019) utilized both the person-object graph and the pose graph of individuals to represent actions between people and objects, as well as interactions among pairs of people. Moreover, GRM (Wang et al., 2018) constructed a knowledge graph comprising persons and objects within an image, while TRGAT (Guo et al., 2023) further considered higher-order constraints for social relations on an image and achieved better results. More recently, GRIT (Yu et al., 2024) leveraged global self-attention mechanisms and graph representation learning to achieve multi-level conditional attention. While these methods have achieved superior performance in terms of accuracy, they still lack the ability to explain their prediction results and exhibit limited generalization capabilities.

2.2 Reinforcement Learning for MLLMs

Recently, works like OpenAI-o1 (Jaech et al., 2024) and DeepSeek-R1 (Guo et al., 2025) have made significant breakthroughs in lifting the reasoning capability of LLMs through reinforcement learning (RL). Subsequent advancements (Shao et al., 2024; Team et al., 2025; Guo et al., 2025) enhance their proficiency in solving complex tasks in chains, including challenging math and coding problems. Building on these models, SocialGPT (Li et al., 2024) combines the reasoning capability of LLMs with the perceptual power of Vision Foundation Models (VFMs), achieving modular, training-free social relation reasoning.

Meanwhile, recent advances in MLLMs have demonstrated a simple approach to visual reasoning, which involves directly querying a multimodal LLM about an image and receiving a response. For these MLLMs, many efforts (Zhou et al., 2025; Liu et al., 2025a; Zhan et al., 2025; Deng et al., 2025a; Peng et al., 2025; Liu et al., 2025b; Yang et al., 2025; Deng et al., 2025b) have applied RL techniques with verifiable reward mechanisms to further boost visual reasoning performance. Specifically, CLS-RL (Li et al., 2025) further improves the performance of MLLMs on visual classification tasks through reinforcement learning methods. Motivated by these prior works, in this paper, we explore the application of RL techniques to train MLLMs specifically for social relationship reasoning tasks.

3 Methodology

In this section, we present our methodology for training Social-R1. Our approach integrates structured prompting with reinforcement learning to enable end-to-end reasoning from visual inputs to relationship predictions. In detail, we first describe our prompting strategy that encourages explicit reasoning before prediction, followed by our reinforcement learning framework based on Group Relative Policy Optimization (GRPO) (Shao et al., 2024).

3.1 Instruction prompt

Following (Shao et al., 2024), we designed a prompt that encourages models to first engage in a thinking process before ultimately delivering the answer. The prompt is designed as:

- {Question} Output the thinking process in <think> </think> and final answer in <answer> </answer> tags, i.e., <think> reasoning process here </think><answer> answer here </answer>

MLLMs can accurately localize the corresponding object in the image, when answering visual questions (Zhang et al., 2025). Inspired by this, we directly place the bounding boxes of the individuals of interest in the query, and following previous works (Zhang et al., 2024; Li et al., 2025) that use MLLMs for image classification to design our question:

- What are the most likely social relationships between bounding box $\{b_i\}$ and $\{b_j\}$? Choose only one from {Class Name}.

where $\{b_i\}$ and $\{b_j\}$ are replaced by the specific bounding boxes of individuals of interest, and {Class Name} is replaced by the candidate label list. This design enables MLLMs to perform perception and reasoning in a single forward pass.

3.2 GRPO-based Training Strategy

After constructing the instruction prompt, we demonstrate how to enhance the model's reasoning ability regarding relationships between people. Firstly, to capture fine-grained visual clues from the input image I—the basis for our reasoning—we employ a pretrained Vision Transformer (ViT) to extract token-level visual features, keeping its weights frozen during the training stage. Next, we enhance the reasoning skills of LLMs, and leverage their extensive social knowledge to predict and interpret the relationships among individuals in the image.

Specifically, to further strengthen the model's social relationship reasoning ability, we adapt Group Relative Policy Optimization (GRPO) (Shao et al., 2024). GRPO has already demonstrated strong potential for enhancing the model's reasoning ability. It uses the average reward of sampled responses as a baseline for computing advantages. The overall process is illustrated in Figure 2. Given an input question q and the extracted visual features of the corresponding image, the policy model $\pi_{\theta_{old}}$ samples a set of responses $\{o_1, o_2, \ldots, o_G\}$,

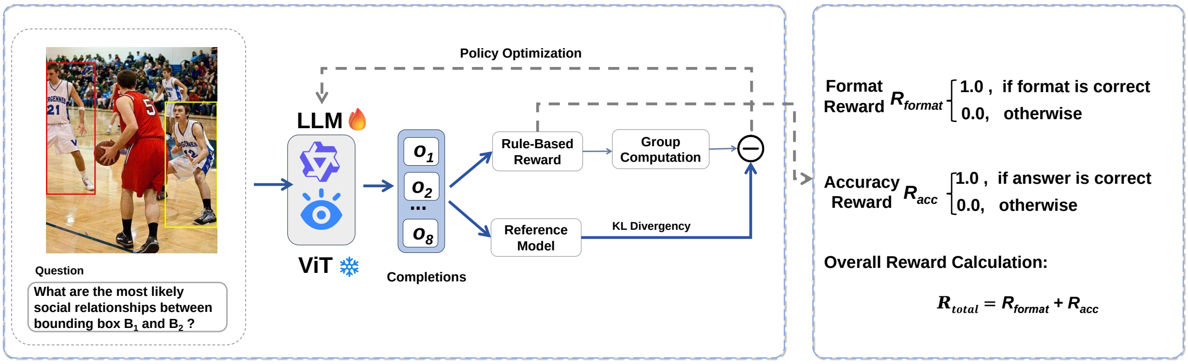


Figure 2: This diagram illustrates a reinforcement learning framework for training vision-language models on social relationship recognition tasks. The left section shows an input image with a question about social relationships between the bounded regions. The center details the policy optimization process, where a vision transformer and language model generate multiple candidate completions that are evaluated through two pathways: a rule-based reward system and a reference model utilizing KL divergence. The right section defines the reward structure, combining format correctness and answer accuracy into a total reward score that guides model optimization.

each evaluated by reward functions to yield scores $\{r_1, r_2, \dots, r_G\}$. The normalized advantage is then computed as:

$$A_i = \frac{r_i - \text{mean}(\{r_1, r_2, \dots, r_G\})}{\text{std}(\{r_1, r_2, \dots, r_G\})} \tag{1}$$

The model π_θ is updated by maximizing the following clipped KL-regularized objective:

$$\begin{aligned} J_{\text{GRPO}}(\theta) &= \mathbb{E}_{\{o_i\}_{i=1}^{G} \sim \pi_{\theta_{\text{old}}}(O|q)} \\ &\Bigg[\frac{1}{G} \sum_{i=1}^{G} \min\Big(\frac{\pi_\theta(o_i \mid q)}{\pi_{\theta_{\text{old}}}(o_i \mid q)} A_i, \\ &\text{clip}\Big(\frac{\pi_\theta(o_i \mid q)}{\pi_{\theta_{\text{old}}}(o_i \mid q)}, 1-\varepsilon,\, 1+\varepsilon \Big) A_i \Big) \\ &- \beta\, D_{\text{KL}}\big(\pi_\theta \,\|\, \pi_{\text{ref}}\big) \Bigg] \end{aligned} \tag{2}$$

The KL divergence loss is computed as:

$$D_{KL}(\pi_\theta \| \pi_{\text{ref}}) = \frac{\pi_{\text{ref}}(o_i|q)}{\pi_\theta(o_i|q)} - \log\left(\frac{\pi_{\text{ref}}(o_i|q)}{\pi_\theta(o_i|q)}\right) - 1 \tag{3}$$

where ε and β are the PPO clipping hyperparameter and the coefficient controlling the Kullback–Leibler (KL) penalty (Shao et al., 2024; Schulman et al., 2017), respectively. GRPO eliminates the critic model in PPO by estimating the relative advantage by sampling a group of responses $\{o_i\}_{i=1}^{G}$ and normalizing their rewards within the group to compute a relative advantage, which is more computationally efficient (Shao et al., 2024).

Reward function. To ensure the reliability and explainability of our results, we employ the Format Reward function. Additionally, to promote the accuracy of the classification results, we utilize the Accuracy Reward function.

1) *Format Reward.* We implement a format-based reward to encourage structured reasoning. The Format Reward guides the model in adopting a standardized response format and optimizing answer selection. Specifically, we require the models to enclose their reasoning process within <think></think> tags and their final answers within <answer></answer> tags.

$$R_{\text{format}} = \begin{cases} 1, & \text{if the format is correct} \\ 0, & \text{if the format is incorrect} \end{cases} \tag{4}$$

2) *Accuracy Reward.* To check the correct answer, we use a regular expression-based approach to evaluate answer accuracy. Specifically, We extract the answer from the <answer> ... </answer> tag, and if it matches the ground-truth answer exactly, it will get a reward of 1 point.

$$R_{\text{answer}} = \begin{cases} 1, & \text{if the answer is correct} \\ 0, & \text{if the answer is incorrect} \end{cases} \tag{5}$$

The total reward, r_i, combines both accuracy and format rewards. Both reward components are indispensable: without the format reward, the final answer cannot be reliably extracted; without the accuracy reward, model training cannot converge as expected.

4 Experiments

4.1 Data and Evaluation

We evaluate our model on two widely-used benchmarks for social relationship reasoning: PIPA (Sun et al., 2017) and PISC (Li et al., 2017). The PIPA dataset defines 16 fine-grained social relationships, including family (e.g., parent-child, grandparent-grandchild), personal (e.g., friends, spouse/lovers), educational and occupational (e.g., teacher-student, leader-subordinate), and group-based (e.g., band, sports team, colleagues) relationships. For evaluation on the PIPA dataset, we use overall accuracy across all categories. Meanwhile, the PISC dataset categorizes social relationships into friend, family, couple, professional, commercial, and no-relation. For the PISC dataset, we report accuracy instead of the mAP (mean Average Precision) metric, because traditional methods calculate mAP based on logit distributions, whereas our model directly outputs text predictions. The data split used in our experiments is summarized in Table 1.

Table 1: Statistics of the PIPA and PISC datasets.

Dataset	Train Pairs	Val Pairs	Test Pairs
PIPA	13,729	709	5106
PISc	55,400	1505	3961

4.2 Experiments settings

Implementation Details. We utilize Qwen2-VL-2B-Instruct (Wang et al., 2024) as the base model. Qwen2-VL (Wang et al., 2024) introduces the Naive Dynamic Resolution mechanism, which enables the model to dynamically process images of varying resolutions into different numbers of visual tokens. We froze the parameters of the visual encoder and fine-tuned the other parameters of the model. The batch size is set to 1 per GPU and we use 2-step gradient accumulation during training. Our implementation builds on public MLLM reasoning toolkits (Chen et al., 2025). For both datasets, the number of candidate generations G is set to 8. We use a learning rate of 2e-6, a temperature parameter of 1, and train each model for 1 epoch. All experiments are conducted on four NVIDIA A6000 GPUs(48 GB). All other hyperparameters are configured according to the settings recommended in Chen et al. (2025).

Baselines. We compare Social-R1 with prior supervised approaches, including Pair CNN (Li et al., 2017), GR^{2}N (Li et al., 2020), GRM (Wang et al., 2018), and GRIT (Yu et al., 2024), which only output answers without providing reasoning rationales. We further compare our method with SocialGPT (Li et al., 2024), an LLM-based method capable of generating rationales, and open-source MLLMs, such as Qwen2-VL-2B and Qwen2-VL-7B, both of which exhibit good performance and interpretability.

Table 2: The comparison results on the PIPA dataset. WR stands for answering with reasoning. WR stands for answering with rationales.

Methods	**WR**	**Acc (%)**
Pair CNN (Li et al., 2017)	✗	58.0
Dual-Glance (Li et al., 2017)	✗	59.6
SRG-GN (Goel et al., 2019)	✗	53.6
GRM (Wang et al., 2018)	✗	62.3
MGR (Zhang et al., 2019)	✗	64.4
GR^{2}N (Li et al., 2020)	✗	64.3
TRGAT (Guo et al., 2023)	✗	65.3
SocialGPT (Li et al., 2024)	✓	66.7
GRIT (Yu et al., 2024)	✗	73.7
Qwen2-VL-2B	✓	64.1
Qwen2-VL-7B	✓	74.0
Social-R1	✓	**80.9**

Table 3: The comparison results on the PISC dataset. Previous methods are replicated with open-source code to report the accuracy metric. WR stands for answering with rationales.

Methods	**WR**	**Acc (%)**
Pair CNN (Li et al., 2017)	✗	46.3
GRM (Wang et al., 2018)	✗	64.2
GR^{2}N (Li et al., 2020)	✗	64.7
SocialGPT (Li et al., 2024)	✓	65.1
GRIT (Yu et al., 2024)	✗	76.0
Qwen2-VL-2B	✓	50.9
Qwen2-VL-7B	✓	55.3
Socia-R1	✓	**76.6**

4.3 Experimental Results

Main Results. Tables 2 and 3 present a comprehensive comparison of our proposed Social-R1 model against previous state-of-the-art approaches on the PIPA and PISC datasets, respectively. The results clearly demonstrate the superior performance of our approach. On the PIPA dataset, Social-R1 achieves an impressive accuracy of 80.9%, significantly outperforming all previous methods. This represents a substantial improvement of 7.2 % over

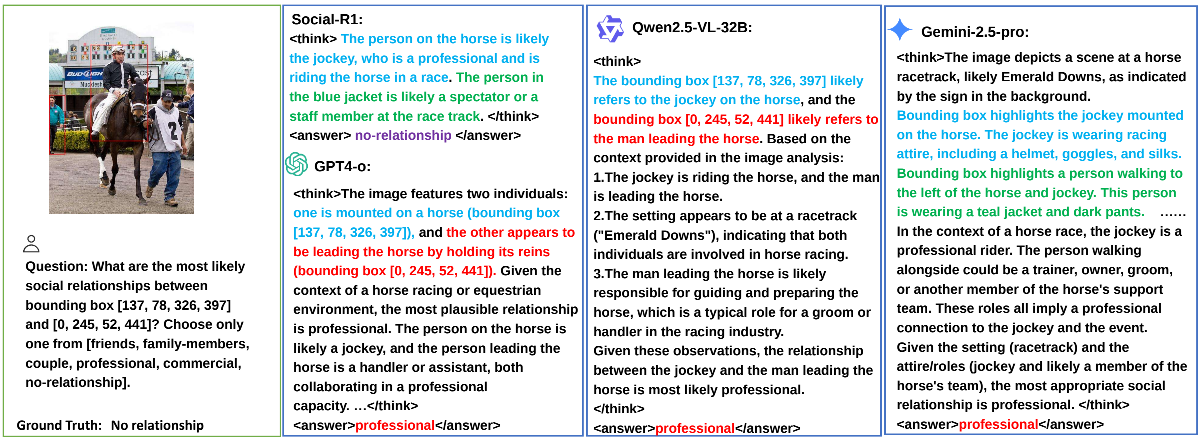


Figure 3: Comparison of baseline models and Social-R1. Explanation of Bias in Various Baseline Models GPT-4o was unable to accurately locate the person in the second bounding box. Qwen2.5-VL-32B cannot describe the content within the second bounding box. Gemini-2.5-pro is able to accurately describe the characteristics of people in both boxes. However, due to the presence of bias, the model still provided an incorrect answer.

Table 4: Comparison with Advanced MLLMs on the PIPA dataset with Social-R1.

Model	WR	Acc(%)
LLaVA	✔	45.1
Qwen2.5-VL-32B	✔	64.9
GPT-4o	✔	75.1
Social-R1	✔	80.9

the previous best model, GRIT (Yu et al., 2024), which achieved 73.7% accuracy. Social-R1 also outperforms Qwen2-VL-7B which achieved 74.0% accuracy, despite being smaller in scale. On the PISC dataset, Social-R1 achieves 76.6% accuracy, surpassing the previous best model GRIT by 0.6%. Compared to the most competitive classification-only methods, we still have improvement.

Notably, among all approaches, only Social-GPT (Li et al., 2024), Qwen2-VL models and our Social-R1 provide answers with explicit reasoning. Compared to SocialGPT, our social-R1 shows significant improvement in accuracy, exceeding by 14.2% on the PIPA dataset and by 11.5% on the PISC dataset. This demonstrates that after reinforcement learning training, our 2B model can outperform SocialGPT's 13B model.

We also noticed that Qwen2-VL-7B performed even better than the best-performing method GTRI on the PIPA dataset. After analyzing its outputs, we discovered that it tends to provide multiple answers within the <answer> tags, as shown in the Figure 6, which led to its inflated accuracy.

Moreover, we also compared Social-R1 with three advanced multimodal large language models, including LLaVA (Liu et al., 2023), Qwen2.5-VL-32B (Bai et al., 2025), and GPT-4o (Hurst et al., 2024) on the PIPA dataset. As results shown in Table 4, we found that, although Social-R1 has a smaller model size, it outperforms the high-performing GPT-4o by 5.8% in accuracyfurther demonstrating the effectiveness of our approach.

4.4 Model Bias

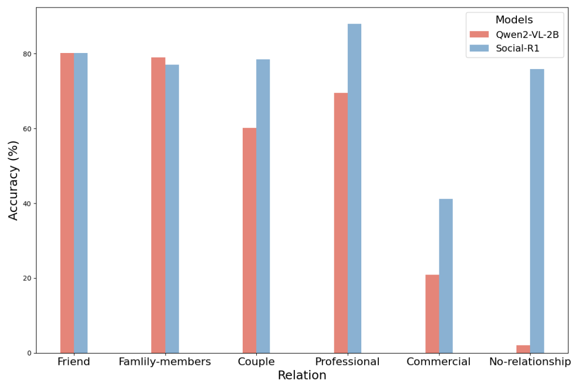


Figure 4: Comparison of relation recognition accuracy between Qwen2-VL-2B and Social-R1 across different relationship categories on the PISC dataset.

Foundation models usually exist biases. As shown in Figure 3, we analyzed the accuracy of the base model Qwen2-VL-2B and Social-R1 on various relationship categories on the PISC dataset. Social-R1 demonstrates consistently higher accu-

Figure 5: Visualization of sample images with attention heatmap (output-to-visual tokens) and rationales.Images in (a) and (b) are from PISC dataset. Images in (c) and (d) are from PIPA dataset.

racy across all categories, with both models performing best in recognizing friendship relations However,the base model had very low accuracy on the "no-relationship" category, only 2.05%, indicating that the foundation model could not identify this relationship. After reinforcement learning, Social-R1 could achieve 75.83% accuracy on the "no-relationship" category.

Moreover, the accuracy of "commercial" is still relatively low after RL training. To investigate this, we studied the composition of the PISC dataset and found a significant data imbalance. As Table 7 shows the Commercial relationship type accounts for only 0.94% (523 samples) of the training set, while the Professional relationship type has the highest proportion at 37.62% (20,842 samples). This imbalanced distribution results in the model receiving significantly insufficient training signals for the "commercial" type during reinforcement learning, thus limiting performance improvement. In contrast, although the Qwen2-VL-2B model initially showed low accuracy on the "no-relationship" category, it achieved significant improvement after training due to abundant training data.

Furthermore, even if the advanced models, exhibit certain biases when performing social relationship classification tasks. Figure 3 is a example that shows bias in baseline model. GPT-4-o (Hurst et al., 2024) exhibited spatial reasoning limitations, failing to accurately locate subjects within the second bounding box. Qwen2.5-VL-32B (Bai et al., 2025) demonstrated a more fundamental constraint in its visual processing capabilities, being unable to describe content within the second bounding box entirely. Interestingly, Gemini-2.5-pro (Team et al., 2024) displayed robust visual perception, accurately describing subject characteristics in both bounding boxes. However, despite this perceptual accuracy, the model still produced biased outputs, indicating that bias persists in its reasoning processes even when visual perception is intact. Social-R1, in contrast, demonstrates improved performance through its novel architecture that better integrates visual understanding with debiased reasoning pathways, as evidenced by our quantitative and qualitative results.

4.5 Ablation Study

Table 5: Ablations on components of Social-R1 on the PIPA datast. Social-R1-vision means training vision encoder during RL training. Social-R1-no-format means training without format reward.

Model	Acc (%)
Social-R1	80.9
Social-R1-vision	80.7
Social-R1-no-format	78.3

To evaluate the contribution of various components within our Social-R1 model, we conducted a series of ablation experiments on the PIPA dataset. Table 5 presents the accuracy results for different model variants. To gain deeper insights into the impact of our design choices, particularly the vision encoder freezing strategy and format reward mechanism, we designed two variant models:

- **Social-R1-vision.** In this variant, we train the vision encoder during the reinforcement learning phase, instead of freezing it. The results show a minimal drop in accuracy to 80.7%, which is 0.2% lower than the baseline model. This suggests that freezing the vision encoder during RL training provides slightly better results, though the difference is marginal.

- **Social-R1-no-format.** This variant removes the format reward mechanism and achieves an accuracy of Social-R1-no-format 78.3%, which is 2.6% lower than the baseline model. This demonstrates that the format reward component contributes substantially to the model's effectiveness. We assume this is because the base model's instruction following ability is relatively weak. The format reward primarily helps the model generate more standardized outputs, improving the normative quality of the model's responses.

4.6 Qualitative evaluation

In this subsection, we visualized the attention weights to illustrate the effectiveness of our method. The specific visualization implementation follows the methodology described in (Zhang et al., 2025). Sample images with attention heatmap and rationales are shown in Figure 5. Our model not only provides the final judgment, but also reveals its "thinking" process – how the model reasons based on visual cues in the image, such as attire, environment, and objects. For example, The clothing of the individuals and the surrounding objects are both described in Figure 5(a) and (b). The attention heatmaps visually demonstrate which areas of the image contributed most to its decision. For instance, the military tank is attended in Figure 5(c), and the track and other individuals are attended in Figure 5(d).

We also present failure cases of Social-R1 in Figure 7. In Figure 7(a), we observe that the model correctly described the people in bounding boxes, however considering the background, the model output the wrong answer "friend" instead of the ground truth "no-relationship". Besides, Social-R1 may lack the ability to attend to age. As illustrated in 7(b), Social-R1 made good descriptions of the image, but it overlooked the the age of the individual, resulting in failure.

5 Conclusion

In this work, we have presented Social-R1, a multimodal large language model trained with reinforcement learning for social relationship reasoning. Our work demonstrates that end-to-end reasoning directly from images and bounding boxes is not only possible but highly effective, eliminating the need for complex multi-stage pipelines or handcrafted prompts. The experimental results on PIPA and PISC benchmarks confirm that our approach achieves state-of-the-art performance while providing interpretable rationales for its predictions. This work contributes to advancing social relationship recognition by combining the reasoning potential of large language models with multimodal capabilities in a more efficient and explainable framework.

Limitations

Despite the strong performance of Social-R1, our work has several limitations that warrant acknowledgment and suggest directions for future research:

Computational Constraints: Due to computational resource limitations, we were unable to train on larger models such as Qwen2-VL-7B. This restricts our understanding of how our approach scales with model size and whether even greater performance improvements could be achieved with larger foundation models.

Reasoning Depth: Although Social-R1 provides explicit reasoning for its predictions, the depth and nuance of this reasoning may still be limited compared to human-level social understanding. The model may rely on visual stereotypes or common patterns rather than deeper contextual reasoning in some cases.

Temporal Dynamics: Our current approach treats social relationship recognition as a static task based on single images, whereas real-world social relationships have temporal dynamics that evolve over time. This static representation limits the model's ability to capture relationship development or changes.

Addressing these limitations presents promising directions for future work, including scaling to larger models, incorporating temporal information, enhancing cross-cultural robustness, and developing more efficient reasoning mechanisms.

References

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. 2025. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*.

Daphne Blunt Bugental. 2000. Acquisition of the algorithms of social life: a domain-based approach. *Psychological bulletin*, 126(2):187.

Liang Chen, Lei Li, Haozhe Zhao, and Yifan Song. 2025. Vinci. r1-v: Reinforcing super generalization ability in vision-language models with less than 3.

Huilin Deng, Ding Zou, Rui Ma, Hongchen Luo, Yang Cao, and Yu Kang. 2025a. Boosting the generalization and reasoning of vision language models with curriculum reinforcement learning. *arXiv preprint arXiv:2503.07065*.

Yihe Deng, Hritik Bansal, Fan Yin, Nanyun Peng, Wei Wang, and Kai-Wei Chang. 2025b. Openvlthinker: An early exploration to complex vision-language reasoning via iterative self-improvement. *arXiv preprint arXiv:2503.17352*.

Alan P Fiske. 1992. The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review*, 99(4):689.

Arushi Goel, Keng Teck Ma, and Cheston Tan. 2019. An end-to-end network for generating social relationship graphs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11186–11195.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Yunfei Guo, Fei Yin, Wei Feng, Xudong Yan, Tao Xue, Shuqi Mei, and Cheng-Lin Liu. 2023. Social relation reasoning based on triangular constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 737–745.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Shinobu Kitayama and Hazel Rose Markus. 2000. The pursuit of happiness and the realization of sympathy: Cultural patterns of self, social relations, and well-being. *Culture and subjective well-being*, 1:113–161.

Junnan Li, Yongkang Wong, Qi Zhao, and Mohan S Kankanhalli. 2017. Dual-glance model for deciphering social relationships. In *Proceedings of the IEEE international conference on computer vision*, pages 2650–2659.

Ming Li, Shitian Zhao, Jike Zhong, Yuxiang Lai, and Kaipeng Zhang. 2025. Cls-rl: Image classification with rule-based reinforcement learning. *arXiv preprint arXiv:2503.16188*.

Wanhua Li, Yueqi Duan, Jiwen Lu, Jianjiang Feng, and Jie Zhou. 2020. Graph-based social relation reasoning. In *European conference on computer vision*, pages 18–34. Springer.

Wanhua Li, Zibin Meng, Jiawei Zhou, Donglai Wei, Chuang Gan, and Hanspeter Pfister. 2024. Socialgpt: Prompting llms for social relation reasoning via greedy segment optimization. *arXiv preprint arXiv:2410.21411*.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916.

Yuqi Liu, Bohao Peng, Zhisheng Zhong, Zihao Yue, Fanbin Lu, Bei Yu, and Jiaya Jia. 2025a. Seg-zero: Reasoning-chain guided segmentation via cognitive reinforcement. *arXiv preprint arXiv:2503.06520*.

Ziyu Liu, Zeyi Sun, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Haodong Duan, Dahua Lin, and Jiaqi Wang. 2025b. Visual-rft: Visual reinforcement fine-tuning. *arXiv preprint arXiv:2503.01785*.

Yingzhe Peng, Gongrui Zhang, Miaosen Zhang, Zhiyuan You, Jie Liu, Qipeng Zhu, Kai Yang, Xingzhong Xu, Xin Geng, and Xu Yang. 2025. Lmm-r1: Empowering 3b lmms with strong reasoning abilities through two-stage rule-based rl. *arXiv preprint arXiv:2503.07536*.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Qianru Sun, Bernt Schiele, and Mario Fritz. 2017. A domain based approach to social relation recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3481–3490.

Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.

Kimi Team, Angang Du, Bofei Gao, Bowei Xing, Changjiu Jiang, Cheng Chen, Cheng Li, Chenjun Xiao, Chenzhuang Du, Chonghua Liao, et al. 2025. Kimi k1. 5: Scaling reinforcement learning with llms. *arXiv preprint arXiv:2501.12599*.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, et al. 2024. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*.

Zhouxia Wang, Tianshui Chen, Jimmy Ren, Weihao Yu, Hui Cheng, and Liang Lin. 2018. Deep reasoning with knowledge graph for social relationship understanding. *arXiv preprint arXiv:1807.00504*.

Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and Philip S Yu. 2023. Multimodal large language models: A survey. In *2023 IEEE International Conference on Big Data (BigData)*, pages 2247–2256. IEEE.

Yi Yang, Xiaoxuan He, Hongkun Pan, Xiyan Jiang, Yan Deng, Xingtao Yang, Haoyu Lu, Dacheng Yin, Fengyun Rao, Minfeng Zhu, et al. 2025. R1-onevision: Advancing generalized multimodal reasoning through cross-modal formalization. *arXiv preprint arXiv:2503.10615*.

Xiaotian Yu, Hanling Yi, Qie Tang, Kun Huang, Wenze Hu, Shiliang Zhang, and Xiaoyu Wang. 2024. Graph-based social relation inference with multi-level conditional attention. *Neural Networks*, 173:106216.

Yufei Zhan, Yousong Zhu, Shurong Zheng, Hongyin Zhao, Fan Yang, Ming Tang, and Jinqiao Wang. 2025. Vision-r1: Evolving human-free alignment in large vision-language models via vision-guided reinforcement learning. *arXiv preprint arXiv:2503.18013*.

Jiarui Zhang, Mahyar Khayatkhoei, Prateek Chhikara, and Filip Ilievski. 2025. Mllms know where to look: Training-free perception of small visual details with multimodal llms. *arXiv preprint arXiv:2502.17422*.

Meng Zhang, Xinchen Liu, Wu Liu, Anfu Zhou, Huadong Ma, and Tao Mei. 2019. Multi-granularity reasoning for social relation recognition from images. In *2019 IEEE international conference on multimedia and expo (ICME)*, pages 1618–1623. IEEE.

Yuhui Zhang, Alyssa Unell, Xiaohan Wang, Dhruba Ghosh, Yuchang Su, Ludwig Schmidt, and Serena Yeung-Levy. 2024. Why are visually-grounded language models bad at image classification? *arXiv preprint arXiv:2405.18415*.

Hengguang Zhou, Xirui Li, Ruochen Wang, Minhao Cheng, Tianyi Zhou, and Cho-Jui Hsieh. 2025. R1-zero's" aha moment" in visual reasoning on a 2b non-sft model. *arXiv preprint arXiv:2503.05132*.

A Statistics of the datasets

The statistics of PIPA and PISC in and fine-grained relations are shown in Tables 6 and 7. Note that the train/val/test splits for PISC coarse and fine are different. We note that both PIPA and PISC have the problem of label imbalance. In particular, in the PISC fine dataset, the relations Couple and Commercial have the smallest numbers of samples among all the relations.

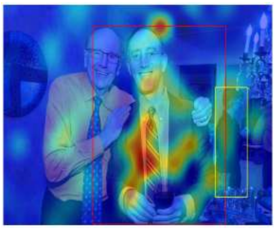
B More Examples

True Label : friends

Qwen2-VL-7B

<think> The image shows two individuals walking together in a natural setting, with one person holding a walking stick. The person holding the walking stick appears to be leading the way, which suggests a guiding or supportive role. The other person is following closely, indicating a level of trust and familiarity. Given the context of hiking, it is likely that these individuals are friends or colleagues who are enjoying a shared outdoor activity. The lack of any formal attire or equipment that would suggest a professional relationship further supports the idea of a casual, social relationship. Therefore, the most likely social relationship between the two individuals is that of friends or colleagues.
</think>
<answer> The most likely social relationship between the bounding box [53, 98, 220, 499] and [210, 49, 332, 414] is that of friends or colleagues. </answer>"

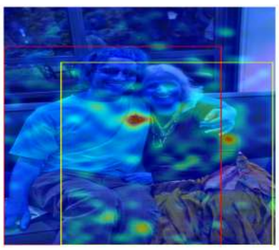
Figure 6: Visualization of Output of Qwen2-VL-7B



(a) True Label : no-relationship

Social-R1

<think> The two men are standing close together, holding a glass of wine, which suggests a social gathering. The fact that they are both wearing ties and suits indicates that they might be at a formal event or a professional gathering. The presence of other people in the background and the setting suggest that they could be friends or colleagues. Therefore, the most likely social relationship between the two men is friends.</think>
<answer> friends </answer> ❌



(b) True Label : mother-child

Social-R1

<think> The person in the yellow shirt is likely the man, and the person in the blue dress is likely the woman. They are sitting together on a couch, which suggests a close relationship, such as friends or family. Therefore, the most likely social relationships are friends or family.</think>
<answer> lovers/spouses </answer> ❌

Figure 7: Visualization of sample images with attention heatmap(output-to-visual tokens) and rationales where Social-R1 makes wrong predictions. Images in (a) and (b) are from PISC and PIPA dataset,respectively

Table 6: Statistics of PIPA fine dataset. We show the number of social relations in train/val/test set.

Category	Train	Valid.	Test
Father–child	332	32	168
Mother–child	448	45	190
Grandpa–grandchild	46	3	11
Grandma–grandchild	37	0	15
Friends	3,054	187	1,833
Siblings	608	32	231
Classmates	128	71	13
Lovers/Spouse	503	49	313
Presenters–audience	194	12	91
Teacher–student	23	15	33
Trainer–trainee	83	1	54
Leader–subordinate	10	1	14
Band members	520	25	211
Dance team members	17	5	326
Sport team members	863	5	294
Colleagues	6,863	226	1,309

Table 7: Statistics of PISC fine dataset. We show the number of social relations in train/val/test set.

Category	Train	Val	Test
Friend	12,686	332	790
Family	7,818	249	677
Couple	1,552	102	256
Professional	20,842	311	858
Commercial	523	164	354
No relation	11,979	347	1,026