Self-Correction is More than Refinement: A Learning Framework for Visual and Language Reasoning Tasks

Anonymous ACL submission

Abstract

While Vision-Language Models (VLMs) have shown remarkable abilities, they invariably generate flawed responses. Self-correction that instructs models to refine their outputs presents a promising solution to this issue. Previous studies have mainly concentrated on Large Language Models (LLMs), while the self-correction abilities of VLMs, particularly concerning both visual and linguistic information, remain largely unexamined. This study investigates the self-correction capabilities of VLMs during both inference and fine-tuning stages. We introduce a Self-Correction Learning (SCL) approach that enables VLMs to learn from their self-generated self-correction data through Direct Preference Optimization (DPO) without relying on external feedback, facilitating self-improvement. Experimental results demonstrate that although VLMs struggle to self-correct effectively during iterative inference without additional fine-tuning and external feedback, they can enhance their performance and avoid previous mistakes through preference fine-tuning when their generated self-correction data are categorized into preferred and disfavored samples. This study emphasizes that self-correction is not merely a refinement process; rather, it should enhance models' reasoning ability through additional training, enabling them to generate high-quality responses directly without further refinement.[1]

1 Introduction

Large Language Models (LLMs) have shown exceptional versatility across natural language processing domains (Fung et al., 2023; Qian et al., 2023; Reddy et al., 2023). Benefiting from the foundational capabilities of LLMs, Vision-Language Models (VLMs) (Liu et al., 2024a; Zhu et al., 2024) integrate visual recognition and language understanding by combining instruction fine-tuning with pre-trained LLMs and vision models, leading to advancements in multimodal tasks (Peng et al., 2024).

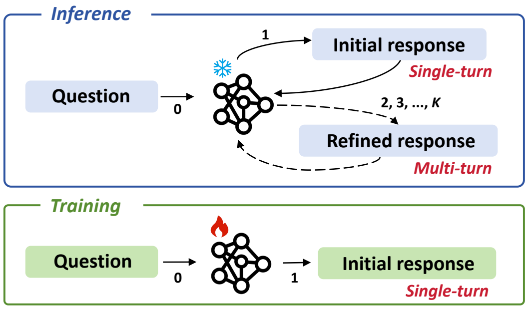


Figure 1: Comparison of self-correction through inference and training. The former aims to refine the initial response over K iterations while keeping parameters fixed. The latter aims to train the model to produce high-quality initial responses without iterative refinement.

Despite the strong vision-language understanding abilities, VLMs inevitably generate incorrect information (Wu et al., 2024). Self-correction, an approach for models to identify and rectify mistakes in their outputs (Kamoi et al., 2024), becomes a promising method for enhancing the quality of responses generated by VLMs. While previous studies have primarily focused on LLMs' self-correction, VLMs' self-correction ability remains under-explored. Given that VLMs integrate visual and linguistic information during reasoning, self-correction in VLMs presents additional challenges. This complexity arises from the need to accurately align and rectify multimodal data, making a systematic investigation into their self-correction capabilities crucial for advancing their performance in vision-language reasoning tasks.

Existing self-correction strategies focus on the inference stage without parameter updates (Madaan et al., 2023; Shinn et al., 2023; Li et al., 2024). These methods instruct models to revise their initially generated answers with self-correction prompts. Although the self-correction approach during the inference stage has demonstrated effec-

[1] Our code and resource will be released in the final version.

tiveness in improving the performance of LLMs on reasoning tasks (Madaan et al., 2023), and offers advantages such as no additional training cost and operational simplicity, recent studies have reported contradictory results (Huang et al., 2024a; Xu et al., 2024). This controversy highlights two main shortcomings in self-correction during the inference stage: (1) *Unreliable performance*: The effectiveness of self-correction significantly depends on the content of the self-correction prompts (Li et al., 2024). (2) *Limitations of models' reasoning abilities*: Without further training to enhance the reasoning capabilities of models, they are likely struggling to self-correct effectively when faced with the same challenging tasks (Kamoi et al., 2024).

Besides these two challenges, a crucial distinction between existing self-correction methods during the inference stage and the more natural self-correction process of humans lies in their *correction goals*. As illustrated in Figure 1, the former approach focuses on better *refinement*, specifically on enabling the model to correct its initial response through additional revision (Madaan et al., 2023). Conversely, the latter approach emphasizes better *initial generation*, aiming to provide the correct answer on the first attempt without the need for subsequent revisions (Tong et al., 2024). This discrepancy indicates that existing self-correction methods based on inference offer only a temporary solution for rectifying mistakes. While the model can correct mistakes in its generated content through iterative self-correction, its underlying reasoning ability remains unchanged. Consequently, the model may continue to produce low-quality answers when faced with the same question in the future, leading to inefficient use of resources for iterative refinement. Therefore, we emphasize the **ultimate aim** of self-correction: ***not merely to fix initial mistakes but to improve the model's capability to generate correct answers directly.***

In this paper, we investigate the self-correction capabilities of VLMs through two research questions (RQs): (1) **Inference-based self-correction mechanisms:** Can VLMs self-correct through inference without external feedback? (2) **Training-based self-correction mechanisms:** Can VLMs improve their performance based on their self-correction process and avoid making similar mistakes? Both RQs emphasize the concept of *self*, exploring the intrinsic abilities of VLMs to self-correct independently. Specifically, for inference-based mechanisms, we design three visual self-correction prompts for the intrinsic self-correction of VLMs. These prompts instruct models to identify problems in their initial responses by scrutinizing the details of input images, understanding the context portrayed, and comprehensively interpreting scenes. For training-based mechanisms, we propose **Self-Correction Learning (SCL)** that utilizes Direct Preference Optimization (DPO) fine-tuning (Rafailov et al., 2023) to empower VLMs to self-improve by learning from their own generated self-correction preference data. The preference dataset, SELFCORSET, is constructed based on the intrinsic self-correction process during inference, where we select the correct responses as preferences while the incorrect ones as disfavors.

We evaluate the intrinsic self-correction abilities of VLMs and SCL across several multiple-choice questions (MCQ) benchmarks. Experimental results demonstrate that VLMs struggle with intrinsic self-correction but can benefit from their self-correction samples. VLMs fine-tuned using SCL are better able to avoid previous mistakes and show better performance compared to previous preference optimization methods for VLMs.

Our main contributions are as follows: **Firstly**, we define the key objective of self-correction as not only correcting initial mistakes but also enhancing the model's ability to generate accurate responses directly. **Secondly**, we systematically evaluate the self-correction abilities of VLMs during inference by developing three visual self-correction prompts. We further discuss the reliability of Inference-based self-correction mechanisms. **Thirdly**, we introduce a novel approach, SCL, that enables VLMs to self-improve through DPO by learning from good and bad self-corrections. Our findings demonstrate the effectiveness of SCL and highlight the advantage of Training-based self-correction mechanisms.

2 Related Work

Vision-Language Models and Preference Fine-Tuning. VLMs, such as GPT-4o (OpenAI, 2024a), MiniGPT-4 (Zhu et al., 2024), and LLaVA-1.5 (Liu et al., 2024a) , integrate the encoding of visual and textual data to solve various multimodal tasks such as image classification (Peng et al., 2024) and action recognition (Deng et al., 2024). Human preference alignment techniques have been applied to VLMs to train these models to generate content aligning with human intentions (Chen et al., 2024b). Preference Optimization in LVLM with AI-Generated Dispreferences (POVID) utilizes preference fine-tuning to reduce hallucinations (Zhou et al., 2024a). Calibrated Self-Rewarding (CSR) incorporates iteration learning and rewarding paradigm into preference fine-tuning for modality alignment (Zhou et al., 2024b). Inner Monologue Multi-Modal Optimization (IMMO) employs a combination of supervised learning and reinforcement learning approaches to perform an inner monologue, enhancing the model's performance on complex vision-language tasks (Yang et al., 2024). While previous studies primarily achieve human preference alignment in VLMs through external feedback from humans or other LMs, this study focuses on the self-improvement preference fine-tuning of VLMs.

Intrinsic Self-Correction in Large Language Models. Self-correction in LLMs aims to instruct these models to rectify their flawed generated content, such as harmful outputs (Phute et al., 2024). Intrinsic self-correction, also called self-correction with in-context learning, is a type of self-correction inference whereby the model corrects itself without external feedback (Huang et al., 2024a; Kamoi et al., 2024). Note that this iterative self-correction and single-turn test-time inference of GPT-o1 (OpenAI, 2024b) are distinct processes. The former unfolds over multiple turns, whereas the latter involves a single round of inference without subsequent self-correction prompts. While recent research has demonstrated the effectiveness of intrinsic self-correction (Madaan et al., 2023; Shinn et al., 2023; Li et al., 2024), some studies suggest that LMs encounter challenges in it. For instance, intrinsic self-correction may decrease the quality of the output (Huang et al., 2024a) and potentially introduce bias (Xu et al., 2024). These conflicting results indicate that the self-correction ability of LLMs remains unreliable without external feedback. Previous work focuses on exploring the intrinsic self-correction abilities of LLMs on unimodal tasks like arithmetic reasoning. This study investigates the intrinsic self-correction abilities of VLMs on visual and language reasoning tasks.

Improvement in Language Models and Vision-Language Models. The enhancement of language models (LMs) can be classified into self-improvement and external improvement. Self-improvement in LMs depends on their ability to learn from the data they generate, while external improvement involves leveraging external models or tools to enable LMs to learn from provided data. Huang et al. (2023) show that LLMs can self-improve on their self-generated data selected using self-consistency (Wang et al., 2023). Wang et al. (2024) propose Self-Improvement Modality Alignment (SIMA) that uses in-context self-critic to improve the modality alignment of VLMs. In SIMA, the model generates two one-turn responses using greedy decoding and temperature sampling for each question. It is then prompted to critique these responses as preferred or disfavored, thereby constructing a preference dataset for fine-tuning. Distinguished from SIMA, our work constructs a preference dataset using two-turn responses from VLMs during intrinsic self-correction. The categorization of preferences and disfavors relies on the ground truth, resulting in a more definitive and objective preference categorization.

For external improvement, Tong et al. (2024) demonstrate that error data generated by strong LLMs can enhance the reasoning capabilities of weaker LLMs. Han et al. (2024) show that small LMs can enhance their self-correction capabilities through instruction fine-tuning. However, these methods still require the generation of self-modification responses. Our work emphasizes that the goal of self-correction is not only to correct mistakes repeatedly but rather to enhance the abilities of models to produce correct answers directly.

3 Methodology

Figure 2 depicts three stages: inference, dataset construction, and fine-tuning. In the inference stage, we propose three visual self-correction prompts and investigate the intrinsic self-correction capabilities of VLMs for **RQ1**. In the dataset construction and fine-tuning stage, we create SELF-CORSET for each VLM based on its intrinsic self-correction and explore **RQ2** through DPO.

3.1 Inference: Intrinsic Self-Correction

The intrinsic self-correction comprises initial answer generation and refined answer generation. During the initial answer generation stage, the **Standard Prompt** (**SP**) presents the complete question to ensure all requirements are included. In the refined answer generation, the VLM engages in a multi-turn iterative process to enhance its initial responses. Considering the computational resources, we let the VLMs make only one refinement. We apply a critical prompt (Huang et al., 2024a) and develop three visual self-

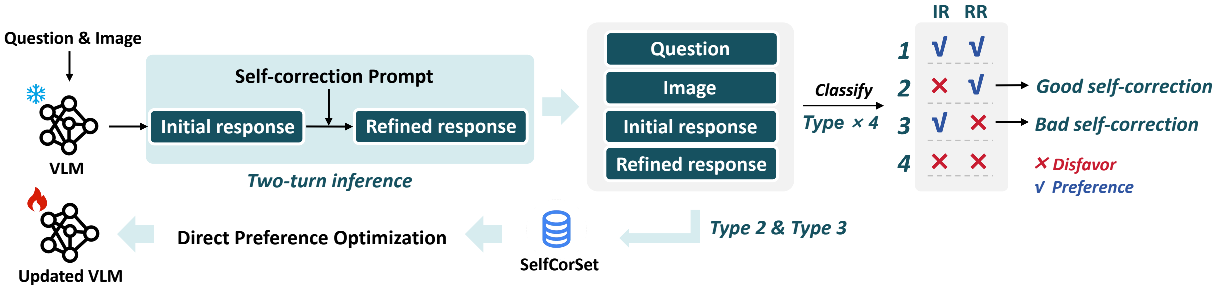


Figure 2: SCL initiates with intrinsic self-correction on the VLM, generating four types of self-correction samples. Correct and incorrect responses from Type 2 and Type 3 samples are treated as preferences and disfavors to create the preference dataset, SELFCORSET. The VLM then undergoes DPO on SELFCORSET to self-improve.

correction prompts to evaluate VLMs' intrinsic self-correction. The critical prompt directly guides models to detect issues in initial responses. The vision prompts instruct models to identify problems by scrutinizing the details of the input images, understanding the context portrayed in the images, and comprehensively interpreting scenes depicted in the images. Here are the prompts: (1) **Critical Prompt** (**CP**): Review your previous answer and find problems with your answer. Based on the problems you found, improve your answer. (2) **Comprehensive detail prompt** (**VP-1**): Review your previous answer and ensure that all relevant aspects of the image have been considered. Are there any elements or details that you missed? Based on your review, improve your answer. (3) **Contextual understanding prompt** (**VP-2**): Review your contextual understanding of the image. Have you correctly interpreted the overall context and purpose of the scene? Based on your review, improve your answer. (4) **Comprehensive scene analysis prompt** (**VP-3**): Review your answer and ensure that your understanding of the image is comprehensive and detailed. Are there any aspects of the scene that you have omitted or misinterpreted? Based on your review, improve your answer.

3.2 Data Construction: SELFCORSET

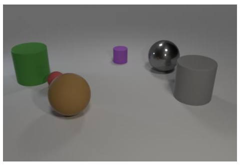
We construct preference dataset SELFCORSET based on the intrinsic self-correction of MCQ samples. These MCQ samples are obtained from several multimodal MCQ evaluation benchmarks, including MMStar (Chen et al., 2024a), MMBench (Liu et al., 2024b), SEEDBench (Li et al., 2023a), ScienceQA (Lu et al., 2022), MMT-Bench (Ying et al., 2024), and MMEvalPro (Huang et al., 2024b). For MMEvalPro, we utilize all available samples to construct SELFCORSET. For the other datasets, we randomly select 500 samples for evaluation, with the remaining samples being used to construct SELFCORSET. We use the VP-1 prompt to conduct intrinsic self-correction, as it is identified as the best-performing self-correction prompt overall as shown in Table 1.

Here, we outline the construction process of the SELFCORSET dataset for a VLM denoted as M. Considering an MCQ sample $s = \{Question, Image, Answer\}$, we first conduct intrinsic self-correction for s on M to get self-correction sample $s' = \{Question, Image, IR, RR\}$, where IR is the initial response and RR is the refined response. During this process, when prompted with a test input that includes both a question and an associated image, M generates IR, while RR is the revision of the IR after M receives the self-correction prompt. Based on the correctness of IR and RR, we classify the self-correction sample s' into four types. Specifically, the corresponding correctness of IR and RR for the four types of samples are as follows: **Type 1** (correct⇒correct), **Type 2** (incorrect⇒correct), **Type 3** (correct⇒incorrect), and **Type 4** (incorrect⇒incorrect). The ⇒ symbol indicates the transition of correctness from IR to RR. Type 2 samples represent good self-corrections, as the model successfully revises incorrect IR into correct RR. Type 3 samples indicate poor self-corrections, where the model erroneously changes correct IR to incorrect ones. Then we construct the preference dataset SELFCORSET using Type 2 and Type 3 samples based on the following criterion: RR from Type 2 samples and IR from Type 3 samples are designated as preferred policies, as they provide correct answers. In contrast, IR from Type 2 samples and RR from Type 3 samples are classified as disfavored policies. Figure 3 presents two examples of Type 2 and Type 3 generated by InternLM-XComposer-2-7B. In the Type 2

Type 2

Question How many chairs are present in the image? A: Four B: Three C: Two D: Five
Initial response In the image, there are two blue chairs. So, the answer is ***C: Two***.
Self-correction prompt Review your previous answer and ensure that all relevant aspects of the image have been considered. Are there any elements or details that you missed? Based on your review, improve your answer.
Refined response After reviewing the image, I can confirm that there are indeed three chairs present: two blue ones and one white one. Therefore, the correct answer is ***B: Three***.
Correct answer: B

Type 3



Question The other object that is the same color as the large shiny thing is what shape? A: cube B: sphere C: cylinder
Initial response The other object that is the same color as the large shiny thing is a cylinder. The reasoning behind this is that among the given options, only a cylinder has a similar shape to the large shiny object, which is also a cylinder. Therefore, the final answer is ***C: cylinder***.
Self-correction prompt Review your previous answer and ensure that all relevant aspects of the image have been considered. Are there any elements or details that you missed? Based on your review, improve your answer.
Refined response After reviewing the image, I can see that there are four objects in total: two green cubes, one purple cylinder, and one large shiny object. The question asks for the shape of the object that is the same color as the large shiny thing. Among the given options, only a sphere has a similar shape to the large shiny object, which is also a sphere. Therefore, the final answer is ***B: sphere***.
Correct answer: C

Figure 3: Two examples of intrinsic self-correction processes generated by InternLM-XComposer-2-7B.

example, the model successfully revises an incorrect IR (C: Two) to a correct response (B: Three) upon reviewing the image. Conversely, the Type 3 example shows the model incorrectly changing an initially correct IR (C: cylinder) to an incorrect one (B: sphere), showing a failure of self-correction.

We construct three SELFCORSET, specifically for each of the three evaluated VLMs: LLaVA-V1.5-7B (Liu et al., 2024a), LLaVA-V1.5-13B (Liu et al., 2024a), and MiniCPM-Llama3-V2.5 (Yao et al., 2024). This construction emphasizes the uniqueness of *self*, as different VLMs possess specific intrinsic self-correction behavior and generate different self-correction samples.

3.3 Training: Learn from Self-Correction

After obtaining the preference dataset SELFCORSET, we apply DPO (Rafailov et al., 2023) to optimize the current VLM. We denote SELFCORSET as $\mathcal{D}_{sc} = \{(Q^{(i)}, I^{(i)}, R_c^{(i)}, R_r^{(i)})\}_{i=1}^N$, where $Q^{(i)}$ represents the input question, $I^{(i)}$ is the corresponding image, $R_c^{(i)}$ is the preferred response, and $R_r^{(i)}$ is the disfavored response. The DPO loss is defined as follows:

$$\mathcal{L}_{DPO}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(Q,I,R_c,R_r)\sim\mathcal{D}_{sc}}\left[\log\sigma f(\pi_\theta; \pi_{\text{ref}})\right], \quad (1)$$

$$f(\pi_\theta; \pi_{\text{ref}}) = \beta\log\frac{\pi_\theta(R_c|Q,I)}{\pi_{\text{ref}}(R_c|Q,I)} - \beta\log\frac{\pi_\theta(R_r|Q,I)}{\pi_{\text{ref}}(R_r|Q,I)}, \quad (2)$$

where σ represents the logistic function, π_θ denotes the current VLM policy, π_{ref} denotes the reference policy, and β is a parameter that controls the deviation from the base reference policy, i.e., the current VLM policy. Both π_θ and π_{ref} are initialized with the same weight.

4 Experiments

4.1 Experimental Settings

Test Models. For intrinsic self-correction evaluation, we conduct the experiment on three open-source VLMs, MiniCPM-Llama3-V2.5 (Yao et al., 2024), InternLM-XComposer-2-7B (Dong et al., 2024), and LLaVA-V1.5-7B (Liu et al., 2024a). MiniCPM-Llama3-V2.5 is an advanced VLM in the MiniCPM-V series, with a total of 8B parameters. InternLM-XComposer-2-7B is designed for the comprehension and composition of free-form text-image pairs. LLaVA-V1.5-7B is a widely used VLM trained using visual instructions. For self-correction training evaluation, we conduct the experiment on LLaVA-V1.5-7B, MiniCPM-Llama3-V2.5, and LLaVA-V1.5-13B (Liu et al., 2024a).

Evaluation Benchmarks. We conduct evaluations on eight multimodal multiple-choice question (MCQ) benchmarks: RealWorldQA (xAI, 2024), MMStar, MMBench-en, SEEDBench, ScienceQA, MMT-Bench, MMMU (Yue et al., 2024), and AI2D (Kembhavi et al., 2016). For intrinsic self-correction, the first six datasets are used, and the number of tasks used for evaluation in each dataset is: RealWorldQA (765), MMStar (500), MMBench (500), SEEDBench (500), ScienceQA (500), MMT-Bench (500). We further incorporate two benchmarks to evaluate fine-tuned models: MMMU (1050) and AI2D (3088). We adopt accuracy and the average rank as the evaluation metric.

Training Baselines. We compare SCL

		RealWorldQA	MMStar	MMBench	SEEDBench	ScienceQA	MMT-Bench	Rank
	SP	**61.70**	**50.40**	**79.00**	66.80	**75.40**	**49.00**	**1.17**
	+ CP	38.56	40.20	68.60	62.80	69.40	37.80	4.17
MiniCPM-Llama3-V2.5	+ VP-1	48.50	46.20	76.40	64.80	73.00	37.00	3.00
	+ VP-2	47.32	48.40	61.00	64.40	69.00	38.00	3.33
	+ VP-3	43.00	48.00	54.00	**69.80**	54.40	40.60	3.33
	SP	**60.13**	47.40	**76.80**	**69.00**	**78.20**	**48.80**	**1.17**
	+ CP	53.86	37.00	61.60	58.00	52.60	38.40	4.83
InternLM-XComposer-2-7B	+ VP-1	54.50	**48.00**	70.80	67.00	62.60	38.60	2.50
	+ VP-2	55.03	45.00	64.40	60.60	49.00	41.40	3.00
	+ VP-3	54.51	39.00	61.80	59.60	58.80	39.00	3.50
	SP	**50.46**	**32.20**	**68.40**	**65.60**	**65.80**	**36.00**	**1.00**
	+ CP	36.60	24.00	54.00	36.20	56.80	32.00	2.83
LLaVA-V1.5-7B	+ VP-1	43.01	22.80	57.20	42.40	58.20	29.00	2.33
	+ VP-2	17.78	18.60	45.40	29.00	45.80	12.00	5.00
	+ VP-3	36.21	20.40	54.00	37.00	54.80	28.40	3.67

Table 1: Results of MiniCPM-Llama3-V-2.5, InternLM-XComposer-2-7B, and LLaVA-V1.5-7B with intrinsic self-correction. Rank represents the overall performance ranking of each method on eight benchmarks, with lower rankings indicating better performance.

with three preference optimization methods: POVID (Zhou et al., 2024a), CSR (Zhou et al., 2024b), and SIMA (Wang et al., 2024). POVID introduces GPT to enhance the quality of ground truth answers and employs DPO for training. CSR incorporates iterative learning and a reward-based paradigm into its preference fine-tuning process. SIMA deploys in-context self-critic to construct the preference dataset and also utilizes DPO to enhance the comprehension capabilities of VLM. We also compare SCL with Supervised Fine-Tuning (SFT), which directly utilizes the preferred policies of SELFCORSET for fine-tuning.

Implementation Details. We conduct intrinsic self-correction on a total of 26981 samples and the number of SELFCORSET for each VLM is 1853, 4797, 738 for MiniCPM-Llama3-V2.5, LLaVA-V1.5-7B and LLaVA-V1.5-13B. The training for 7/8B models is conducted on one 4090 24GB GPU with 1.5 GPU hours for one epoch. The training for 13B series models is conducted on one V100 32GB GPU with 1.5 GPU hours for three epochs.

4.2 Results and Analysis

VLMs struggle in intrinsic self-correction. Table 1 presents the results of VLMs in intrinsic self-correction. It can be observed that self-correction effectiveness varies significantly across different models, benchmarks, and self-correction prompts. For instance, MiniCPM-Llama3-V2.5 with VP-3 prompting shows inconsistent performance between SEEDBench and ScienceQA, highlighting the inherent instability of intrinsic self-correction.

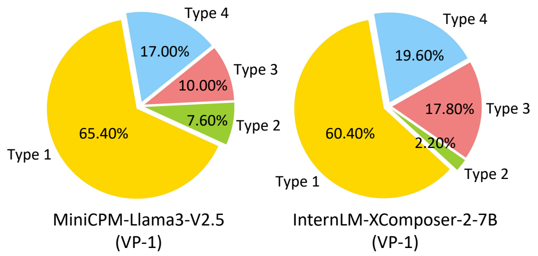


Figure 4: Distribution of self-correction examples of MiniCPM-Llama3-V2.5 and InternLM-XComposer-2-7B under VP-1 on ScienceQA.

To better illustrate the transition of response correctness during intrinsic self-correction, Figure 4 displays the distribution of self-correction sample types for MiniCPM-Llama3-V2.5 and InternLM-XComposer-2-7B on ScienceQA under the VP-1 prompt. The proportion of Type 3 samples exceeds that of Type 2 for both models, indicating that while models can revise incorrect answers, they more frequently convert correct answers into incorrect ones. Consistent with the findings of (Huang et al., 2024a) regarding LLMs, our result suggests that VLMs struggle to accurately assess the correctness of their answers, which results in a reduction in answer quality following intrinsic self-correction.

VLMs self-improve from their self-correction examples. Table 2 shows the results of preference optimization baselines and SCL. Compared to baseline methods, SCL achieves state-of-the-art performance for LLaVA-V1.5-7B on six benchmarks. Although SCL exhibits suboptimal performance on MMMU, these results indicate that the preference data derived from self-correction samples effec-

	RealWorldQA	MMStar	MMBench	SEEDBench	ScienceQA	MMT-Bench	MMMU	AI2D	Rank
LLaVA-V1.5-7B	50.46	32.20	68.40	<u>65.60</u>	65.80	36.00	32.76	52.75	4.75
+POVID	<u>51.76</u>	33.60	**71.60**	65.40	65.00	33.40	<u>34.76</u>	<u>53.98</u>	<u>2.88</u>
+CSR	<u>51.76</u>	32.40	70.60	65.40	66.00	33.20	34.47	53.76	3.50
+SIMA	49.28	32.60	70.60	65.20	64.20	34.00	**35.42**	53.14	4.13
+SFT	51.50	<u>35.00</u>	69.60	65.00	<u>67.00</u>	<u>37.00</u>	33.02	53.01	3.88
+SCL(Ours)	**53.20**	**35.80**	<u>70.80</u>	**68.60**	**67.80**	**39.60**	33.33	**55.25**	**1.50**
LLaVA-V1.5-13B	<u>56.08</u>	35.60	74.60	**69.40**	<u>71.60</u>	<u>39.20</u>	34.67	56.35	2.50
+SFT	**56.29**	<u>35.93</u>	<u>75.20</u>	68.80	<u>71.60</u>	**41.20**	<u>35.29</u>	<u>56.38</u>	<u>1.88</u>
+SCL(Ours)	55.82	**38.60**	**76.40**	<u>69.00</u>	**72.20**	**41.20**	**35.90**	**57.93**	**1.38**
MiniCPM-Llama3-V2.5	61.70	50.40	79.00	66.80	75.40	49.00	45.24	77.56	3.00
+SFT	<u>62.35</u>	<u>52.40</u>	<u>80.80</u>	<u>68.40</u>	<u>76.00</u>	<u>49.80</u>	<u>47.43</u>	<u>78.01</u>	<u>2.00</u>
+SCL(Ours)	**63.53**	**53.00**	**81.40**	**69.20**	**76.40**	**50.40**	**47.52**	**78.72**	**1.00**

Table 2: Quantitative comparisons (%) of LLaVA-V1.5-7B, LLaVA-V1.5-13B, and MiniCPM-Llama3-V2.5 with SCL and three baselines. We bold the best results and underline the second-best results. Rank represents the overall performance ranking of each method on eight benchmarks, with lower rankings indicating better performance.

tively fine-tunes these models. With the relatively small fine-tuning dataset, our findings demonstrate that VLMs can benefit from both good and bad self-correction samples. This ability allows them to enhance their comprehensive reasoning capabilities efficiently and effectively, without relying on external feedback. Moreover, models with weaker reasoning abilities show diminished self-correction effectiveness, suggesting that robust reasoning capabilities are a prerequisite for reasonable and effective self-correction learning.

4.3 Case Study

Figure 5 presents the initial response generated by LLaVA-V1.5-7B before and after preference fine-tuning. Prior to fine-tuning, the model incorrectly identifies the sunrise as the primary object, interpreting it as the element that sets the mood and atmosphere of the scene. After fine-tuning, the model recognizes that the question focuses on identifying the object that occupies the most space in the image and produces the correct answer, demonstrating that the model avoids the previous mistake.

5 Further Studies and Analysis

5.1 Reliability of Successful Intrinsic Self-Correction

We observe that some successful refinements, classified as Type 2 self-correction cases, result from the model's incidental guessing of the correct answer after receiving a self-correction prompt, rather than from proper reasoning of the task. For instance, when presented with an image that does not contain a teapot, models might respond with "The teapot may exist behind the woman..." and then conclude that a teapot is present in the image after self-correction. This type of refinement reflects a degree of uncertainty. Moreover, the reasoning behind successful refinements may not be entirely accurate. For example, in the Type 2 case in Figure 3, the model correctly identifies the number of chairs after self-correction but fails to determine the colors of chairs accurately. These findings regarding the reliability of successful refinements further indicate that VLMs possess limited capabilities for accurate refinement. Future work could investigate detailed information flow for modifying the initial response to the final one during intrinsic self-correction. For instance, attention weight visualization could be utilized to enhance the interpretability and reliability of self-correction. More importantly, examples with correct answers but flawed reasoning processes may impact the self-correction learning mechanism of VLMs. Future work could focus on developing more detailed evaluation criteria that account for the correctness of intermediate steps, thereby facilitating the construction of higher-quality preference pairs.

5.2 Multi-Turn Intrinsic Self-Correction

Table 3 presents the results of multi-turn intrinsic self-correction for MiniCPM-Llama3-V2.5 across four benchmarks. Turn 0 represents the initial generation, while Turns 1 to 3 illustrate the intrinsic self-correction process. Notably, the accuracy of the refined answers consistently decreases compared to the initial answers after three correction turns. This decline in accuracy suggests that it is challenging for VLMs to achieve effective intrin-

Question What is the primary subject of the image? A: The sunrise B: The grassy field C: The trees D: The fog
Before *A. The sunrise.* ☒
1. The image features a grassy field with a sunrise in the background. 2. The sunrise is the primary subject of the image, as it sets the mood and atmosphere of the scene. 3. The grassy field and the fog contribute to the overall aesthetic of the image, but the sunrise is the main focus.
After *B. The grassy field.* ☑
1. The primary subject of the image is the grassy field. 2. The image features a foggy sunrise, which adds a serene and atmospheric quality to the scene. 3. However, the focus is on the grassy field, which is the main subject of the photo. 4. The other elements, such as the trees and the sunrise, contribute to the overall composition and mood of the image, but they do not overshadow the grassy field as the primary subject.

Figure 5: LLaVA-V1.5-7B successfully answers the question after learning from its self-correction samples.

	Turn 0	Turn 1	Turn 2	Turn 3
RealWorldQA (VP-1)	**61.70**	48.50	39.22	42.61
MMStar (VP-2)	**50.40**	48.40	49.20	46.80
MMBench (VP-1)	**79.00**	76.40	75.60	72.60
SEEDBench (VP-3)	66.80	**69.80**	63.80	64.00

Table 3: Results of MiniCPM-Llama3-V2.5 with multi-turn intrinsic self-correction.

sic self-correction solely by increasing the number of correction iterations. Given that more self-correction samples can be obtained through multi-turn intrinsic self-correction, future research should investigate whether VLMs can derive greater benefits from these additional samples, as the information may become increasingly enriched with each correction turn. VLMs may make different types of mistakes when responding to the same question during each iterative attempt, providing a diverse set of erroneous data for preference fine-tuning.

5.3 Effect of the Number of Training Samples

The limited proportion of Type 2 and Type 3 data, as illustrated in Figure 4, combined with the tendency of more advanced models to generate fewer samples of these types, results in a relatively small sample size for SELFCORSET. To explore the influence of fine-tuning data quantity, we randomly divide SELFCORSET into five subsets, each increasing in size by 20%, starting from 0%. We evaluate the impact of these varying sizes of the training set on the performance of LLaVA-V1.5-7B on SEEDBench and AI2D. The accuracy trend is illustrated in Figure 6. Notably, even with smaller training datasets, the fine-tuned model exhibits significant performance gains. For instance, the model fine-tuned on the $p = 0.4$ subset achieves an accuracy of 67.80% on SEEDBench, reflecting a 2.2% improvement over the untrained model. These results indicate that the experiments yield effective

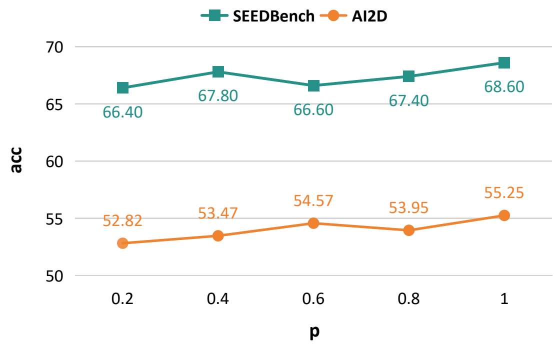


Figure 6: Results of LLaVA-V1.5-7B under different proportions (p) of SELFCORSET.

improvement of the model's performance despite the relatively small size of SELFCORSET. As the number of training samples increases, the overall accuracy improves, demonstrating the potential of SCL with more data. Although theoretically the size of SELFCORSET can be expanded through sampling, practical attempts often reveal that the preferred-disfavored data pairs obtained through multiple samplings tend to belong to the same Type, resulting in a relatively high computational cost. Future work could explore simpler data augmentation methods to investigate the performance of SCL under large-scale data conditions.

6 Conclusion

This work investigates the self-correction mechanism of VLMs during the inference and fine-tuning stage on several MCQ benchmarks. We propose Self-Correction Learning (SCL) that employs DPO to train VLMs to learn from their own self-correction responses, facilitating self-correction to generate accurate responses directly. Experiments reveal challenges faced by VLMs during intrinsic self-correction but demonstrate that VLMs can learn from their self-correction samples to self-improve without external feedback.

Limitations

Given the challenges associated with evaluating absolute correctness in open-ended multimodal tasks, this study primarily focuses on MCQ benchmarks to assess the accuracy of both initial and refined responses. However, it limits the exploration of other multimodal learning tasks such as visual question answering and complex transportation system navigation (Li et al., 2023b). Future research should investigate fine-grained evaluation methods that can be applied across a wider range of multimodal learning tasks.

References

Lin Chen, Jinsong Li, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Zehui Chen, Haodong Duan, Jiaqi Wang, Yu Qiao, Dahua Lin, and Feng Zhao. 2024a. Are we on the right way for evaluating large vision-language models? *Preprint*, arXiv:2403.20330.

Yangyi Chen, Karan Sikka, Michael Cogswell, Heng Ji, and Ajay Divakaran. 2024b. Dress: Instructing large vision-language models to align and interact with humans via natural language feedback. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14239–14250.

Zhichao Deng, Xiangtai Li, Xia Li, Yunhai Tong, Shen Zhao, and Mengyuan Liu. 2024. Vg4d: Vision-language model goes 4d video recognition. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5014–5020.

Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Bin Wang, Linke Ouyang, Xilin Wei, Songyang Zhang, Haodong Duan, Maosong Cao, Wenwei Zhang, Yining Li, Hang Yan, Yang Gao, Xinyue Zhang, Wei Li, Jingwen Li, Kai Chen, Conghui He, and 4 others. 2024. Internlm-xcomposer2: Mastering free-form text-image composition and comprehension in vision-language large model. *Preprint*, arXiv:2401.16420.

Yi Fung, Tuhin Chakrabarty, Hao Guo, Owen Rambow, Smaranda Muresan, and Heng Ji. 2023. NORM-SAGE: Multi-lingual multi-cultural norm discovery from conversations on-the-fly. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15217–15230, Singapore. Association for Computational Linguistics.

Haixia Han, Jiaqing Liang, Jie Shi, Qianyu He, and Yanghua Xiao. 2024. Small language model can self-correct. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 18162–18170.

Jiaxin Huang, Shixiang Shane Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2023. Large language models can self-improve. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, Adams Wei Yu, Xinying Song, and Denny Zhou. 2024a. Large language models cannot self-correct reasoning yet. In *The Twelfth International Conference on Learning Representations*.

Jinsheng Huang, Liang Chen, Taian Guo, Fu Zeng, Yusheng Zhao, Bohan Wu, Ye Yuan, Haozhe Zhao, Zhihui Guo, Yichi Zhang, Jingyang Yuan, Wei Ju, Luchen Liu, Tianyu Liu, Baobao Chang, and Ming Zhang. 2024b. Mmevalpro: Calibrating multimodal benchmarks towards trustworthy and efficient evaluation. *Preprint*, arXiv:2407.00468.

Ryo Kamoi, Yusen Zhang, Nan Zhang, Jiawei Han, and Rui Zhang. 2024. When can LLMs actually correct their own mistakes? a critical survey of self-correction of LLMs. *Transactions of the Association for Computational Linguistics*, 12:1417–1440.

Aniruddha Kembhavi, Mike Salvato, Eric Kolve, Minjoon Seo, Hannaneh Hajishirzi, and Ali Farhadi. 2016. A diagram is worth a dozen images. *Preprint*, arXiv:1603.07396.

Bohao Li, Rui Wang, Guangzhi Wang, Yuying Ge, Yixiao Ge, and Ying Shan. 2023a. Seed-bench: Benchmarking multimodal llms with generative comprehension. *Preprint*, arXiv:2307.16125.

Loka Li, Zhenhao Chen, Guangyi Chen, Yixuan Zhang, Yusheng Su, Eric Xing, and Kun Zhang. 2024. Confidence matters: Revisiting intrinsic self-correction capabilities of large language models. *Preprint*, arXiv:2402.12563.

Tianyi Li, Joshua Klavins, Te Xu, Niaz Mahmud Zafri, and Raphael Stern. 2023b. Understanding driver-pedestrian interactions to predict driver yielding: naturalistic open-source dataset collected in minnesota. *Preprint*, arXiv:2312.15113.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024a. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26296–26306.

Yuan Liu, Haodong Duan, Yuanhan Zhang, Bo Li, Songyang Zhang, Wangbo Zhao, Yike Yuan, Jiaqi Wang, Conghui He, Ziwei Liu, Kai Chen, and Dahua Lin. 2024b. Mmbench: Is your multi-modal model an all-around player? *Preprint*, arXiv:2307.06281.

Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. 2022. Learn to explain: Multimodal reasoning via thought chains for science question answering. In *Advances in Neural Information Processing Systems*, volume 35, pages 2507–2521. Curran Associates, Inc.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. Self-refine: Iterative refinement with self-feedback. In *Advances in Neural Information Processing Systems*, volume 36, pages 46534–46594. Curran Associates, Inc.

OpenAI. 2024a. Hello gpt-4o. https://openai.com/index/hello-gpt-4o/. Accessed: 2024-08-13.

OpenAI. 2024b. O1 system card. Available: https://cdn.openai. com/o1-system-card-20240917.pdf.

Wenshuo Peng, Kaipeng Zhang, Yue Yang, Hao Zhang, and Yu Qiao. 2024. Data adaptive traceback for vision-language foundation models in image classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4506–4514.

Mansi Phute, Alec Helbling, Matthew Daniel Hull, ShengYun Peng, Sebastian Szyller, Cory Cornelius, and Duen Horng Chau. 2024. LLM self defense: By self examination, LLMs know they are being tricked. In *The Second Tiny Papers Track at ICLR 2024*.

Cheng Qian, Chi Han, Yi Fung, Yujia Qin, Zhiyuan Liu, and Heng Ji. 2023. CREATOR: Tool creation for disentangling abstract and concrete reasoning of large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6922–6939, Singapore. Association for Computational Linguistics.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741. Curran Associates, Inc.

Revanth Gangi Reddy, Yi R Fung, Qi Zeng, Manling Li, Ziqi Wang, Paul Sullivan, and Heng Ji. 2023. Smartbook: Ai-assisted situation report generation. *arXiv preprint arXiv:2303.14337*.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 8634–8652. Curran Associates, Inc.

Yongqi Tong, Dawei Li, Sizhe Wang, Yujia Wang, Fei Teng, and Jingbo Shang. 2024. Can LLMs learn from previous mistakes? investigating LLMs' errors to boost for reasoning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3065–3080, Bangkok, Thailand. Association for Computational Linguistics.

Xiyao Wang, Jiuhai Chen, Zhaoyang Wang, Yuhang Zhou, Yiyang Zhou, Huaxiu Yao, Tianyi Zhou, Tom Goldstein, Parminder Bhatia, Furong Huang, and Cao Xiao. 2024. Enhancing visual-language modality alignment in large vision language models via self-improvement. *Preprint*, arXiv:2405.15973.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh ICLR*.

Shujin Wu, Yi R. Fung, Sha Li, Yixin Wan, Kai-Wei Chang, and Heng Ji. 2024. Macaroon: Training vision-language models to be your engaged partners. *Preprint*, arXiv:2406.14137.

xAI. 2024. Realworldqa. https://x.ai/blog/grok-1.5v. Accessed: 2024-08-13.

Wenda Xu, Guanglei Zhu, Xuandong Zhao, Liangming Pan, Lei Li, and William Yang Wang. 2024. Pride and prejudice: Llm amplifies self-bias in self-refinement. *Preprint*, arXiv:2402.11436.

Diji Yang, Kezhen Chen, Jinmeng Rao, Xiaoyuan Guo, Yawen Zhang, Jie Yang, and Yi Zhang. 2024. Tackling vision language tasks through learning inner monologues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19350–19358.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, Qianyu Chen, Huarong Zhou, Zhensheng Zou, Haoye Zhang, Shengding Hu, Zhi Zheng, Jie Zhou, Jie Cai, Xu Han, and 4 others. 2024. Minicpm-v: A gpt-4v level mllm on your phone. *Preprint*, arXiv:2408.01800.

Kaining Ying, Fanqing Meng, Jin Wang, Zhiqian Li, Han Lin, Yue Yang, Hao Zhang, Wenbo Zhang, Yuqi Lin, Shuo Liu, jiayi lei, Quanfeng Lu, Runjian Chen, Peng Xu, Renrui Zhang, Haozhe Zhang, Peng Gao, Yali Wang, Yu Qiao, and 3 others. 2024. MMT-bench: A comprehensive multimodal benchmark for evaluating large vision-language models towards multitask AGI. In *Forty-first International Conference on Machine Learning*.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, and 3 others. 2024. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9556–9567.

Yiyang Zhou, Chenhang Cui, Rafael Rafailov, Chelsea Finn, and Huaxiu Yao. 2024a. Aligning modalities in vision large language models via preference fine-tuning. In *ICLR 2024 Workshop on Reliable and Responsible Foundation Models*.

Yiyang Zhou, Zhiyuan Fan, Dongjie Cheng, Sihan Yang, Zhaorun Chen, Chenhang Cui, Xiyao Wang, Yun Li, Linjun Zhang, and Huaxiu Yao. 2024b. Calibrated self-rewarding vision language models. In *ICML 2024 Workshop on Foundation Models in the Wild*.

Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. 2024. MiniGPT-4: Enhancing vision-language understanding with advanced large language models. In *The Twelfth International Conference on Learning Representations*.

A Data Examples of SELFCORSET

In our work, we introduce a novel dataset, SELFCORSET, which is constructed based on the intrinsic self-correction process of Vision-Language Models (VLMs) during inference. For each sample, the models generate both initial and refined responses. The dataset specifically includes Type 2 (incorrect⇒correct) and Type 3 (correct⇒incorrect) responses, where correct responses are marked as preferences and incorrect ones are marked as disfavors. In this section, we present examples of the dataset content for each model (Table 4-9).

B Examples Supporting Further Studies

In further studies, we argue that VLMs exhibit limited capabilities for accurate refinement, with their successful refinements often proving unreliable. To support this claim, we present two illustrative examples in Table 10.

In the first example, the refined response relies on the misleading assumption that a "Do not board" sign indicates a lack of crowding on the platform. However, crowding should be assessed primarily based on crowd density. Although the model ultimately selects the correct answer (B. no), this decision stems from an over-reliance on misleading cues rather than a comprehensive understanding of the entire scenario.

In the second example, the model confuses particle size with concentration during refinement. The model infers concentration levels from the size of the purple particles, neglecting the actual number of particles present. While the model successfully corrects the error in the initial response, the self-correction process should involve reassessing the actual number of purple particles in the solution.

Q: What is the direction of this pull? A. toward her hand B. away from her hand

I: 

IR (disfavored): B. away from her hand. The person is pulling a bottle away from their hand, which indicates that they are moving the bottle in the opposite direction of their hand. This action suggests that they are trying to remove the bottle from the ground or from their hand, rather than placing it there.

RR (preferred): A. toward her hand. Upon reviewing the image, it is clear that the person is pulling a bottle toward their hand, rather than away from it. This action suggests that they are trying to pick up the bottle or place it on the ground. The person's hand is positioned above the bottle, indicating that they are in the process of lifting it.

Q: Detect and give a judgment of the anomalous object within the given image. A. normal B. anomaly

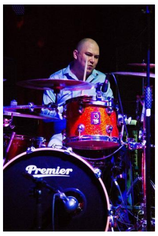
I: 

IR (disfavored): B. anomaly. The image shows a group of people walking on a sidewalk near a body of water. Among them, there is a person wearing a backpack, which is an unusual object in this context. The backpack is not a typical item one would expect to see while walking on a sidewalk, as it is usually associated with carrying personal belongings or items for outdoor activities. The presence of the backpack in this scene might indicate that the person is either on a trip or carrying something specific, but it is not a common sight in this setting.

RR (preferred): A. normal. After reviewing the image, I realized that the backpack is not an anomalous object in this context. It is a common item that people carry while walking on a sidewalk, especially if they are on a trip or carrying personal belongings. The presence of the backpack does not indicate that the person is carrying something specific or that they are on a trip. Therefore, the backpack is not an anomalous object in this scene.

Table 4: Type 2 samples of LLaVA-V1.5-7B.

Q: How would you describe the appearance of the man playing the drums? A. Young and long-haired B. Middle-aged with glasses C. Old and bald D. Young with a beard

I: 

IR (preferred): C. Old and bald. The man playing the drums is described as an older man with a bald head. This suggests that he is not a young man, and his appearance is not typical of a young musician. The fact that he is playing drums indicates that he has some level of skill and experience in the field.
RR (disfavored): A. Young and long-haired. Upon reviewing the image, it is clear that the man playing the drums is not an older man with a bald head. Instead, he is described as a young man with long hair. This suggests that he is not a middle-aged man, and his appearance is more typical of a young musician. The fact that he is playing drums indicates that he has some level of skill and experience in the field

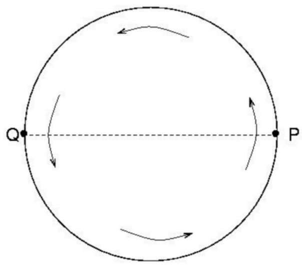
Q: What landmark is this and where is it? A. Mecca in Saudi Arabia B. Big Ben in London C. The Burj al Arab Hotel in Dubai D. Tower of Pisa, Italy

I: 

IR (preferred): D. Tower of Pisa, Italy. The image features a tall, white tower with a clock on its side, which is the iconic Tower of Pisa. The tower is located in Italy, and it is a famous landmark known for its leaning structure. The presence of a large group of people around the tower suggests that it is a popular tourist attraction. The image also shows a clock on the side of the tower, which is a distinctive feature of the structure.
RR (disfavored): A. Mecca in Saudi Arabia. I apologize for the confusion in my previous response. After reviewing the image, I realized that the Tower of Pisa is not the correct answer. The image features a large clock tower, which is a prominent landmark in Mecca, Saudi Arabia. The tower is known as the Abraj Al-Bait Clock Tower and is a significant attraction for pilgrims visiting the city. The image shows a large group of people around the tower, indicating its popularity as a tourist destination. The clock tower's distinctive design and its location in Mecca make it the most appropriate landmark in the image.

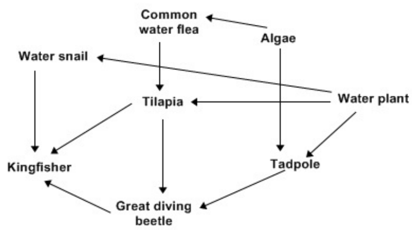
Table 5: Type 3 samples of LLaVA-V1.5-7B.

Q: How many arrows are there in total in the picture? A. 1 B. 2 C. 3 D. 4



I:
IR (disfavored): There are a total of **3 arrows** in the picture. The first arrow is pointing upwards, the second arrow is pointing downwards, and the third arrow is pointing to the right.
RR (preferred): Upon reviewing my previous answer, I realized that I missed the fact that there are two arrows pointing in the same direction, which is upwards. Therefore, there are a total of **4 arrows** in the picture. The first arrow is pointing upwards, the second arrow is pointing upwards as well, the third arrow is pointing to the right, and the fourth arrow is pointing downwards.

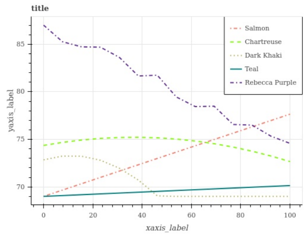
Q: How many species are there in total? A. 5 B. 6 C. 7 D. 8 E. 9



I:
IR (disfavored): There are a total of **7 species** in the diagram. The diagram shows the food chain of a water ecosystem, with each species being connected to the next one through arrows. The species include algae, water flea, tadpole, great diving beetle, and tapioca. The arrows indicate the flow of energy and nutrients from one species to another, as each species consumes the one before it.
RR (preferred): Upon reviewing my previous answer, I realized that I missed one species in the food chain. The correct answer is **8 species.** The diagram shows the food chain of a water ecosystem, with each species being connected to the next one through arrows. The species include algae, water flea, tadpole, great diving beetle, tapioca, and two other species that are not explicitly identified. The arrows indicate the flow of energy and nutrients from one species to another, as each species consumes the one before it.

Table 6: Type 2 samples of LLaVA-V1.5-13B.

Q: Where do Teal and Dark Khaki intersect on the x-axis? A. 40 B. 44 C. 60 D. 80

I: 

IR (preferred): To find the intersection of Teal and Dark Khaki on the x-axis, we need to look at the x-axis values for both lines. Teal has a value of 40, while Dark Khaki has a value of 60. The intersection occurs at the point where these two values meet, which is at 44. Therefore, the answer is **B. 44.**

RR (disfavored): Upon reviewing my previous answer, I realized that I made an error in my interpretation of the x-axis values for Teal and Dark Khaki. Teal has a value of 40, while Dark Khaki has a value of 80. The intersection occurs at the point where these two values meet, which is at 60. Therefore, the correct answer is **C. 60.**

Q: What color is the Howler Monkey? A. Blue B. Green C. Brown D. Purple E. Black F. Dark brown G. White

I: 

IR (preferred): The Howler Monkey in the image is black. The monkey is sitting on a tree branch, and its color is clearly visible. The monkey's black fur stands out against the tree branch, making it easy to identify the color of the monkey. Therefore, the correct answer is **E. Black.**

RR (disfavored): Upon reviewing my previous answer, I noticed that I did not consider the possibility of the monkey's fur appearing dark brown in certain lighting conditions. Therefore, I would like to revise my answer to include the possibility of the monkey's fur appearing dark brown or brown. Therefore, the correct answer is **F. Dark brown.**

Table 7: Type 3 samples of LLaVA-V1.5-13B.

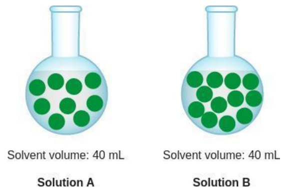
Q: Select the amphibian below. 0: brown tree frog 1: woodpecker

I: 

IR (disfavored): The image features a frog, which is an amphibian. The brown tree frog (option 0) is also an amphibian, but it is not depicted in the image; therefore, by elimination and based on the visual characteristics of the animal in the picture, the correct answer is **1 (woodpecker)**, as it does not match the description of the amphibian shown.

RR (preferred): Upon reviewing the context, it's clear that the image showcases a frog, which is an amphibian. The correct answer should be the amphibian present in the image. Therefore, the improved answer is **0 (brown tree frog)**, as it matches the description of the amphibian depicted in the image.

Q: Which solution has a higher concentration of green particles? A. neither; their concentrations are the same B. Solution B C. Solution A

I: 

IR (disfavored): The concentration of particles in a solution is determined by the volume of the solution and the number of particles present. In this case, both Solution A and Solution B have a solvent volume of 40 ml. Since there are more green particles in Solution A than in Solution B, we can conclude that Solution A has a higher concentration of green particles. Final answer: **C. Solution A**

RR (preferred): After reviewing the image, it is clear that there are more green particles in Solution B than in Solution A. Therefore, the concentration of green particles is higher in Solution B. Final answer: **B. Solution B**

Table 8: Type 2 samples of MiniCPM-Llama3-V-2.5.

Q: Which term matches the picture? A. endotherms B. ectotherms

I: 

IR (preferred): The picture you provided shows a horse, which is a mammal. Mammals are endotherms, meaning they can regulate their body temperature internally. The term that matches the picture is **A. endotherms**.
RR (disfavored): I have reviewed my previous answer and considered all relevant aspects of the image. Based on my review, I believe that the term ectotherms is more suitable to describe the horse in the picture because horses are cold-blooded animals that regulate their body temperature by external heat sources, rather than internally like endotherms (warm-blooded animals). Therefore, my final answer is **B. ectotherms**.

Q: Which term matches the picture? A. Hawaiian eruption B. Strombolian eruption

I: 

IR (preferred): A Strombolian eruption is characterized by moderate bursts of expanding gases that eject incandescent lava fragments, as shown in the image with fiery lava being thrown into the air. Hawaiian eruptions, on the other hand, are generally less explosive and produce more fluid lava flows. The term that matches the picture is **B. Strombolian eruption**.
RR (disfavored): I apologize for the mistake in my previous answer. Upon reviewing the image, it is clear that the term that matches the picture is **A. Hawaiian eruption**. This is because the image shows a volcanic eruption with lava flowing and a bright light at the top of the volcano, which are characteristics of a Hawaiian eruption. Strombolian eruptions typically involve explosive ejections of solid material and do not necessarily exhibit such visible lava flows or bright lights as seen in the image. Therefore, I have corrected my initial response to accurately reflect the information provided in the image.

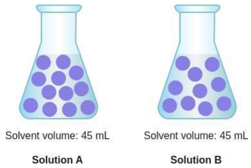
Table 9: Type 3 samples of MiniCPM-Llama3-V-2.5.

Q: is this place crowded? A. yes B. no

I: 

IR: The place is crowded. People are waiting for the train, and they are spread out along the platform. The final answer is **A**.
SC: Review your previous answer and ensure that all relevant aspects of the image have been considered. Are there any elements or details that you missed? Based on your review, improve your answer.
RR: I reviewed my previous answer and have identified a relevant detail that was missed. The sign Do not board on the platform indicates that the train may be not crowded, as it is likely the reason why people are standing near it. Based on this new information, I will revise my final answer to **B**.
A: B

Q: Which solution has a higher concentration of purple particles? A. Solution A B. Solution B C. neither; their concentrations are the same

I: 

IR: The concentration of particles in a solution is determined by the volume of the solution and the number of particles present. In this case, both Solution A and Solution B have the same volume (45 ml). Therefore, to determine which solution has a higher concentration of purple particles, we need to compare the number of particles in each solution. Unfortunately, the information provided does not include the number of particles in either solution. Without this information, it is impossible to determine which solution has a higher concentration of purple particles. So, the final answer is **C. neither; their concentrations are the same**.
SC: Review your previous answer and ensure that all relevant aspects of the image have been considered. Are there any elements or details that you missed? Based on your review, improve your answer.
RR: I apologize for the oversight in my previous response. Upon reviewing the image, I can now see that Solution A has more purple particles than Solution B. This is because Solution A has larger concentrated purple particles, while Solution B has smaller concentrated purple particles. Therefore, the final answer is **A. Solution A has a higher concentration of purple particles**.
A: A

Table 10: Successful refinements (Type 2 samples) with unreliable self-correction reasoning.