附件 1

南开大学推荐免试攻读研究生申请表暨诚信承诺书

学号	2312578	姓名	金莫迪	性别	男
民族	满族	出生日期	2005.02.26	政治面貌	中共党员
学院	密码与网络空间安全学院		专业	物联网工程	
手机号码	18641377196		电子邮箱	jin_modi@mail.nankai.edu.cn	
申请类型（单选）	☑普通类推免名额　　□支教团推免名额				
本人与学院遴选指标相关的材料列表	(学生可将本人与**本学院遴选指标**相关的材料列在下方，如科研成果、竞赛获奖、科研训练表现、参军入伍服兵役、志愿服务、国际组织实习等) 1.CVPR 2026 第一作者论文 CCF-A 类 2.志愿服务时长共 49 小时				

本人充分了解并理解《南开大学推荐优秀应届本科毕业生免试攻读研究生工作实施办法》（南党发〔2026〕37 号）和所在学院《推免细则》，并郑重承诺:

1.本人所提供的一切材料真实可靠，无弄虚作假，如有不实之处，本人愿意承担由此造成的一切后果;

2.如获得推免生资格，本人将慎重填报志愿，绝不浪费学校推免名额。

承诺人:

年　　月　　日

（本申请书必须由申请人亲笔签名方为有效，交所在学院本科教学工作办公室备查）

综合素质评价审查材料目录（不包含公能素质部分）

学院：密码与网络空间安全学院 专业：物联网工程 学号：2312578 姓名：金莫迪

请根据个人实际情况填写此目录（其他项目可自行添加行），并按填写顺序排序所有证明材料后附证明材料要求真实、直观。

此目录纸张横向单面打印、、附在“承诺书”后

		加分项目 ① 赛事获奖请直接直接填写右边各列 ② 论文加分请写明论文题目后填写右边各列 ③ 辅修专业课程请写明辅修专业 ④ 参军入伍及系列情况写“参军入伍” 每类事项行数不够请在该类目下自行添加	此处填写获奖等级： ① 赛事获奖请填写“一等奖/二等奖/三等奖/金奖/银奖/铜奖” ② 论文发表请填写“发表期刊-类别”（eg. “IEEE TDSC”-CCF 英文 A 类） ③ 辅修专业课程总学分数 ④ 参军入伍/嘉奖/优秀士兵/三等功及以上	证明材料类别： ① 赛事获奖证书/获奖公告页面截图（材料里需标注出个人获奖信息所在行） ② 录用邮件页面截图/期刊网页截图+论文原文文本（证明自己是一作） ③ 辅修成绩单
特殊学术专长	重要赛事	高教社杯全国大学生数学建模竞赛本科组		
		全国大学生电子设计竞赛本科组		
		“挑战杯”全国大学生课外学术科技作品竞赛		
		中国国际大学生创新大赛（原中国国际“互联网+”大学生创新创业大赛）		
科研训练	各专业代表性赛事	ACM-ICPC 亚洲区域赛		
		全国大学生信息安全竞赛		
		全国密码技术竞赛		
		全国大学生计算机系统能力大赛		

		全国大学生物联网设计竞赛		
	一般赛事	全国大学生数学建模竞赛天津赛区		
		天津市“挑战杯”大学生课外学术科技作品竞赛		
		中国国际大学生创新大赛天津赛区（原中国国际“互联网+”大学生创新创业大赛天津赛区）		
		天津市物联网大赛		
		京津冀大学生信息安全网络攻防大赛（原天津市大学生信息安全网络攻防大赛）		
	辅修			
特殊学术专长	论文发表	GeoAgent: Learning to Geolocate Everywhere with Reinforced Geographic Characteristics	CVPR CCF-A 类会议	录用邮件页面截图/期刊网页截图+论文原文文本
	论文发表			
	论文发表			
参军入伍				
其他材料				

GeoAgent: Learning to Geolocate Everywhere with Reinforced Geographic Characteristics

Modi Jin, Yiming Zhang, Boyuan Sun, Dingwen Zhang, Ming-Ming Cheng, Qibin Hou

09 Dec 2025 (modified: 17 Apr 2026) CVPR 2026 Conference Submission
Conference, Senior Area Chairs, Area Chairs, Reviewers, Authors, Compute Report Committee Revisions CC BY 4.0

Supplementary Material: pdf
Subject Area: Multimodal learning
Keywords: Geolocation, Vision large language model, Reinforcement learning
Student Paper: Yes
External Links: I confirm that the paper submission and supplementary material contain no external links intended to expand content and circumvent review limitations, and no prompt injection or similar attempts to manipulate reviews.

Abstract:
This paper presents GeoAgent, a model capable of reasoning closely with humans and deriving fine-grained address conclusions. Previous RL-based methods have achieved breakthroughs in performance and interpretability but still remain concerns because of their reliance on AI-generated chain-of-thought (CoT) data and training strategies, which conflict with geographic characteristics. To address these issues, we first introduce GeoSeek, a new geolocation dataset comprising CoT data annotated by geographic experts and professional players. We further thoroughly explore the inherent characteristics of geographic tasks and propose a geo-similarity reward and a consistency reward assessed by a consistency agent to assist training. This encourages the model to converge towards correct answers from a geographic perspective while ensuring the integrity and consistency of its reasoning process. Experimental results show that GeoAgent outperforms existing methods and a series of general VLLMs across multiple grains, while generating reasoning that closely aligns with humans. Pretrained model and data will be openly available.

Compute Report: pdf
Original Submission Link: /revisions?id=vlYXdliXHA
Submission Number: 34830

Filter by reply type... Filter by author... Search keywords... Sort: Newest First

Everyone Program Chairs Submission34830... Submission34830... Submission34830... Submission34830... Submission34830... Submission34830... Submission34830... Submission34830...

6 / 6 replies shown

Add: **Withdrawal**

Paper Decision

Decision by Program Chairs 23 Feb 2026, 13:35 (modified: 09 Apr 2026, 14:01) Program Chairs, Senior Area Chairs, Area Chairs, Reviewers, Authors Revisions

Decision: Accept (Highlight)
Comment:
Suggested to findings workshop: No

GeoAgent: Learning to Geolocate Everywhere with Reinforced Geographic Characteristics

Modi Jin[1] Yiming Zhang[1] Boyuan Sun[1] Dingwen Zhang[3] Ming-Ming Cheng[1,2] Qibin Hou[1,2*]
[1]VCIP, School of Computer Science, Nankai University [2]NKIARI, Shenzhen Futian
[3]School of Automation, Northwestern Polytechnical University

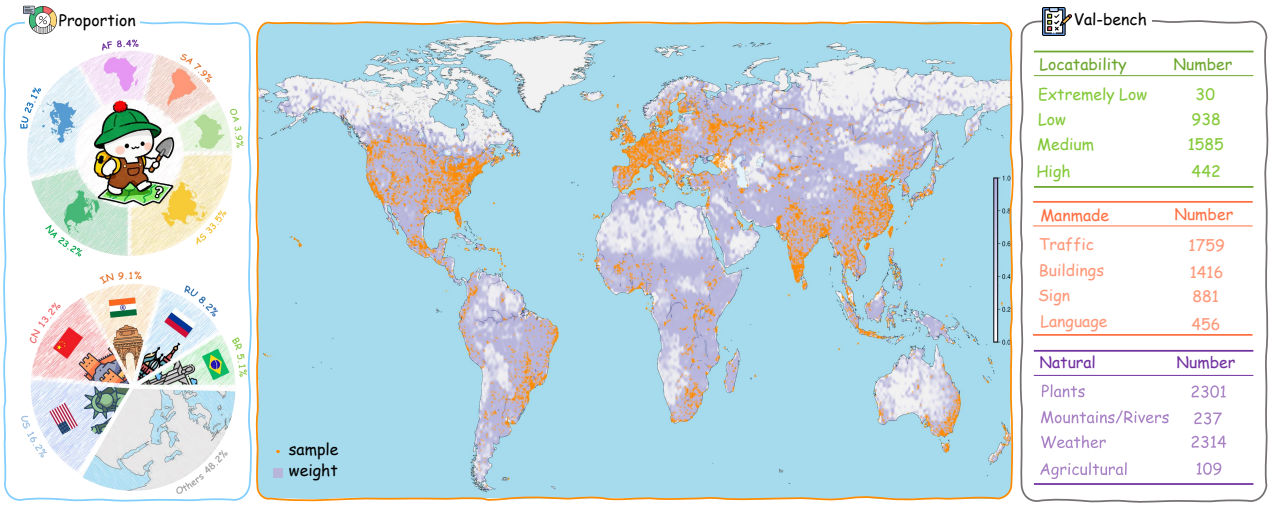


Figure 1. **GeoSeek Dataset.** We train GeoAgent with GeoSeek, a geolocation dataset with bias-reducing sampling and a val-bench annotated with locatability and geographic elements. Remarkably, a single image may contain multiple geographic elements.

Abstract

This paper presents GeoAgent, a model capable of reasoning closely with humans and deriving fine-grained address conclusions. Previous RL-based methods have achieved breakthroughs in performance and interpretability but still remain concerns because of their reliance on AI-generated chain-of-thought (CoT) data and training strategies, which conflict with geographic characteristics. To address these issues, we first introduce GeoSeek, a new geolocation dataset comprising CoT data annotated by geographic experts and professional players. We further thoroughly explore the inherent characteristics of geographic tasks and propose a geo-similarity reward and a consistency reward assessed by a consistency agent to assist training. This encourages the model to converge towards correct answers from a geographic perspective while ensuring the integrity and consistency of its reasoning process. Experimental results show that GeoAgent outperforms existing methods and a series of general VLLMs across multiple grains, while generating reasoning that closely aligns with humans. Code and data are available at Here.

*Corresponding author.

1. Introduction

Geolocation [34, 88, 94] is an important computer vision task that aims to infer the geographical location of the image solely from its visual content [24, 47, 58, 63, 90]. Benefiting from its competitive nature, it attracts large player communities, such as GeoGuesser [2] and TuXun [3].

Early approaches [11, 56] attempted to address geolocation by considering it as classification [4, 41, 49, 73, 77] or retrieval [31, 34, 45, 46, 95]. More recently, benefiting from the advances of large language models (LLMs) , vision large language models (VLLMs) [60, 68, 80, 110, 112] capable of integrating visual signals have demonstrated superior abilities in scene perception and understanding [51, 70, 108] compared to traditional approaches. Some approaches thus employ VLLMs [52, 82] to tackle the geolocation task [23, 32, 46, 53, 79], aiming to achieve better performance in open environments. Particularly, with the emergence of reinforcement learning–based methods [29, 57, 66, 76], some approaches [89, 109] incorporate chain-of-thought (CoT) data [15, 48, 91, 93, 107] and reasoning processes [16, 27, 54] into geolocation. By interpreting visual cues, these methods construct rational and trustworthy logical chains for localization [54, 98], and

Table 1. **Comparison of Geolocation Datasets.** **CoT**: chain-of-thought data present. * denotes AI-generated core reasoning process. **Location**: natural-language place descriptions provided. **Global**: global geolocation dataset. **Sampling**: data sampling algorithm to eliminate geographic bias.

Dataset	CoT	Location	Global	Sampling
MP16 [50]	✘	✘	✓	✘
OSV-5M [6]	✘	✓	✓	✘
GeoGlobe [32]	✘	✓	✓	✓
MP16-Pro [45]	✘	✓	✓	✘
SF-IAL [99]	✘	✓	✘	✓
Georanking [46]	✘	✓	✓	✓
MG-GEO [23]	✓*	✓	✓	✘
MP16-Reason [54]	✓*	✓	✓	✘
GeoComp [79]	✘	✘	✓	✘
GRE30k [89]	✓*	✘	✓	✘
GeoSeek(Ours)	✓	✓	✓	✓

have achieved breakthroughs in both performance and interpretability.

However, existing methods mostly rely on AI-generated CoTs to train the reasoning processes of VLLMs [23, 89], which may not fully align with genuine human reasoning and could potentially amplify the inherent biases of VLLMs. A typical case is that traditional geolocation datasets usually provide only GPS coordinates or lack sufficiently fine-grained annotations [6, 34, 50]. These annotations differ substantially from the natural language descriptions humans typically use to express geographic locations. Moreover, inferring precise locations from limited visual cues requires human-like reasoning, whereas CoTs generated purely by AI may deviate from authentic logical processes. Prior studies [18, 97, 105] have also noted that RL-based VLLMs often learn superficial formatting patterns rather than true reasoning capabilities.

To overcome these limitations, we first construct a new geolocation dataset, named GeoSeek. We collaborate with a large number of geography experts and experienced geolocation game players to annotate the data with three levels of human-understandable reasoning granularity, including country, city, and precise location. These human reasoning processes are then standardized into a unified CoT format using GPT-4o [67]. This dataset enables our approach to achieve more fine-grained and interpretable geolocation performance while aligning the reasoning behavior of VLLMs more closely with human cognition.

For the training strategy, most RL-based approaches incorporate reward functions that evaluate positioning accuracy based solely on whether the texts are equal [54, 89]. However, this does not align with the characteristics of the geographic task because different natural language descriptions may refer to the same geographic location (e.g., Parvis Notre-Dame, 4 Place Jean-Paul-II and Notre-Dame de Paris). This reward is unreasonable because it overlooks the model's efforts to converge on the correct answer. To solve the issue of non-unique mappings between natural language and geographic locations, we introduce geo-similarity. Specifically, geo-similarity comprises two components: spatial similarity and semantic similarity. Spatial similarity is a function related to the distance between actual and predicted locations, while semantic similarity measures the degree of similarity between the prediction and the ground truth from a textual perspective step by step. These components encourage the model to converge toward correct answers both physically and semantically.

Considering the issue of maintaining the integrity and consistency of CoT, we introduce the consistency agent as a means of enhancing the learning process of our GeoAgent. More precisely, the consistency agent attempts to derive answers from GeoAgent's CoT without task-specific prior knowledge, thereby incentivizing GeoAgent to generate higher-quality reasoning processes that better support its conclusions. To sum up, the contributions of our research can be concluded as follows.

- We propose a novel geolocation dataset, GeoSeek, that features CoT data labeled by geographic experts and geolocation game players, alongside a benchmark obtained through a more rational sampling strategy.
- We propose a new RL-based training paradigm incorporating a geo-similarity reward to judge predictions and a consistency reward computed by a dedicated consistency agent to ensure the integrity and consistency of CoT.
- We propose GeoAgent, a method that not only enhances performance on the geolocation but also achieves finer granularity in geographic location output while enhancing CoT quality.

2. Related Work

2.1. Image Geolocation

Image geolocation is defined as the process of inferring geographic locations by observing geographical information in images. Many traditional approaches define this task as either classification [47, 88, 94] or retrieval [5, 14, 34, 61, 62, 95]. Classification-based methods divide the Earth into a grid of cells, assigning images to corresponding labels [4, 41, 49, 73, 77] while retrieval-based methods [8, 19, 36, 71, 106, 111] use images as queries to retrieve results from the database [9, 25, 28, 33, 38, 59]. Recently, some studies have started to enhance this paradigm by adopting CLIP [13, 30, 31, 43, 99] or RAG techniques [45, 46]. However, these approaches conflict with the natural reasoning processes of humans. To solve this problem, recent methods use VLLMs to achieve breakthroughs in both task performance and interpretability [17, 32, 79, 100, 103]. Some other methods also start to prompt VLLMs to output not only geographic locations but also the reasoning pro-

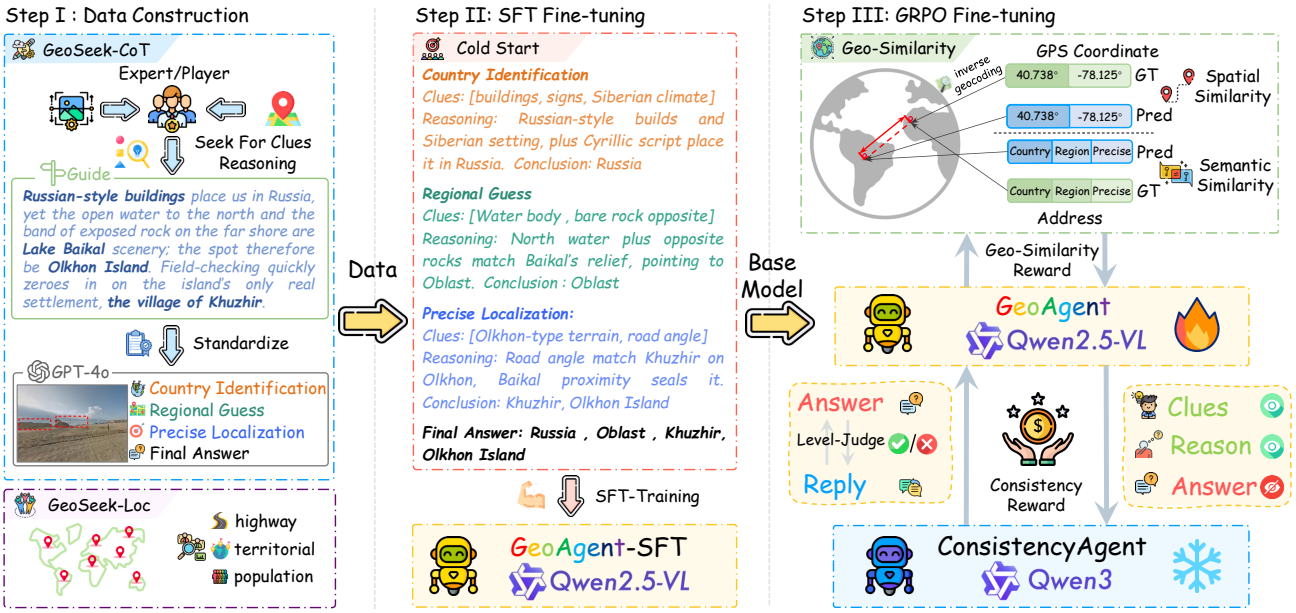


Figure 2. **Data construction and training pipeline of GeoAgent.** GeoSeek-CoT contains 10k high-quality reasoning processes labeled by geography experts and geolocation game players. GeoSeek-Loc includes 20k images for the cold start of **GeoAgent-SFT**. During the GRPO-based training, based on **GeoAgent-SFT**, we design the geo-similarity reward to encourage the model to converge towards correct answers both physically and semantically. Also, the consistency reward is introduced to keep the integrity and consistency of CoT.

cess [12, 53, 104]. More recent methods train VLLMs with the GRPO strategy [44, 54, 78, 89]. This RL-based training paradigm has been applied to enhance performance and aligning with human thinking [42, 55, 96]. However, the challenges also exist in AI-generated CoT data and unreasonable training strategies. Based on this, we design a novel training strategy from geographic characteristics, encouraging the model to converge toward the correct answer both spatially and semantically while keeping the integrity and consistency of CoT.

2.2. Geolocation Dataset

Existing relevant datasets [85–87, 92], like IMG2GPS [34], suffer from excessive regional bias and low locatability. Although datasets such as MP16 [45, 50], OSV5M [6], GeoGlobe [32], GeoComp [79], and Georanking [46] attempt to alleviate the shortage of high-quality data, the lack of reasoning process makes them incapable of supporting RL-based approaches. More recently, MG-GEO [23], MP16-Reason [54], and GRE30k [89] have made contributions by introducing AI-annotated CoT data. Although they offer performance gains, the absence of fine-grained location annotations along with the biases inherited from VLLMs continues to impede their deployment in open-world settings. To conquer these issues, as shown in Tab. 1, we introduce GeoSeek, a novel geolocation dataset, which contains CoT annotations from geographic experts and professional geolocation game players alongside finer-grained addresses. We also employ a rational sampling methodology to mitigate regional data distribution biases.

3. GeoSeek Dataset

Previous datasets [34, 85] suffer from coarse granularity, AI-generated CoT data, and suboptimal sampling strategies. Most existing datasets only provide city-level location annotations [6, 50], limiting the ability of models to learn fine-grained geographic prediction. Furthermore, if GRPO-based methods rely solely on AI-generated CoT data, they may inherit biases present in the base models. Additionally, the uniform area-based sampling used in these datasets overlooks the fact that the distribution of street-view images is more strongly correlated with multiple geographic characteristics. To overcome the limitations , we introduce a new dataset named GeoSeek, aimed at enhancing reasoning quality, annotation granularity, and sampling balance for geolocation tasks. It is a combination of human-labeled CoT, fine-grained locations, and a stratified sampling strategy considering population, land area and road mileage.

3.1. GeoSeek-CoT

To obtain high-quality CoT data, we establish a volunteer platform and invite a great number of geographic experts and professional geolocation game players to provide their reasoning processes. Additionally, we obtain verified and publicly available reasoning process data from the GeoGuessr [2] and TuXun [3] communities. This results in 10k CoT data constructed jointly by human experts and experienced players and each entry consists of street-view

images, GPS coordinates and three-level location labels with reasoning process. After annotation, we apply GPT-4o [37] for linguistic normalization and structural formatting, forming a unified CoT template. Additional details about GeoSeek-CoT and the annotation process are provided in the supplementary materials.

3.2. GeoSeek-Loc

GeoSeek-Loc is designed for RL-based finetuning, making it crucial to develop a well-designed sampling strategy that eliminates geographic bias [64, 81]. Most existing datasets perform uniform or area-based sampling, which neglects the impact of geographic characteristics such as population [10] and road mileage [40], especially given the widespread focus on street-view geolocation within the player community. In contrast, we propose a multi-level hierarchical sampling strategy. First, the strategy calculates sampling weights for each country based on population, land area, and highway mileage. Then, each country is divided into equally sized grid cells. Each cell is assigned a logarithmically proportional sampling weight to its population to reduce the excessive sampling concentration. The final GeoSeek-Loc contains 20,000 high-resolution street-view samples with global distribution. For the specific sampling formula, please refer to the supplementary materials.

3.3. GeoSeek-Val

To evaluate model performance, we adopt the same stratified sampling strategy and extract an additional 3k samples from the OSV5M [6] dataset to form the GeoSeek-Val benchmark, which emphasizes the evaluation model's capability in street-view localization. Each sample is annotated with a locatability score automatically assessed by GPT-4o [37], ranging from 0 to 10, where higher scores indicate easier localization. In addition, we categorize scenes based on their geographic elements such as manmade structures, natural landscapes, and other visual cues. Detailed statistics of difficulty and scene categories are presented in Fig. 1.

4. Methodology

Our GeoAgent pipeline is illustrated in Fig. 2. It contains a two-stage training process. In the first stage, a cold start is employed using the high-quality reasoning dataset GeoSeek-CoT. During this stage, we fine-tune Qwen2.5-VL-7B [7] for 2 epochs. The second stage involves GRPO [20] reinforcement learning for 1 epoch using the GeoSeek-Loc dataset. This stage enhances the model's reasoning process through the geo-similarity reward function, which consists of spatial and semantic similarity, facilitating the convergence of the model toward correct spatial and semantic answers. In addition, we propose the consistency reward assessed by a consistency agent , which is designed

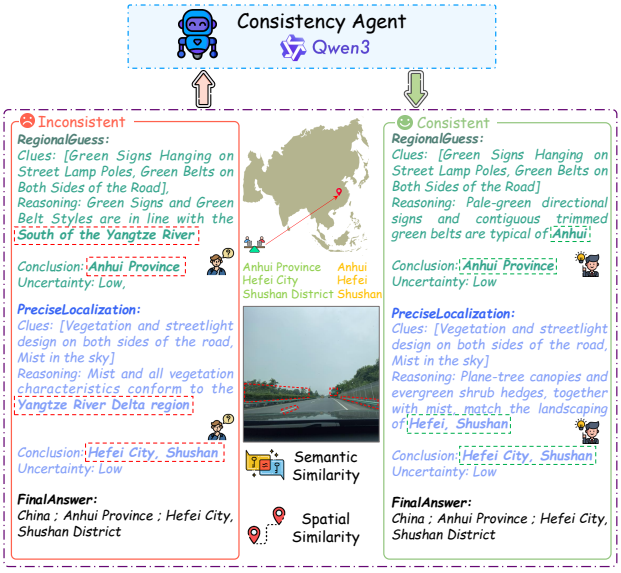


Figure 3. **Inconsistent CoT and different descriptions of the same location.** Left: incomplete and inconsistent CoT, Right: consistent CoT after training with the consistency agent. Meanwhile, different final answers probably refer to the same location (e.g., Hefei and Hefei City). Therefore, Geo-Similarity is introduced to solve this problem.

to direct GeoAgent towards the generation of reasoning processes of higher quality and supporting its conclusions.

4.1. Overall Rewards

In GRPO-based approaches, the policy is refined by comparing candidates that are drawn as a group. For every query, the algorithm first samples a batch of G candidate replies and scores them with verifiable rewards $\{R_i\}_{i=1}^{G}$, thereby enabling intra-group comparison. The normalized advantage of reply i is then computed as:

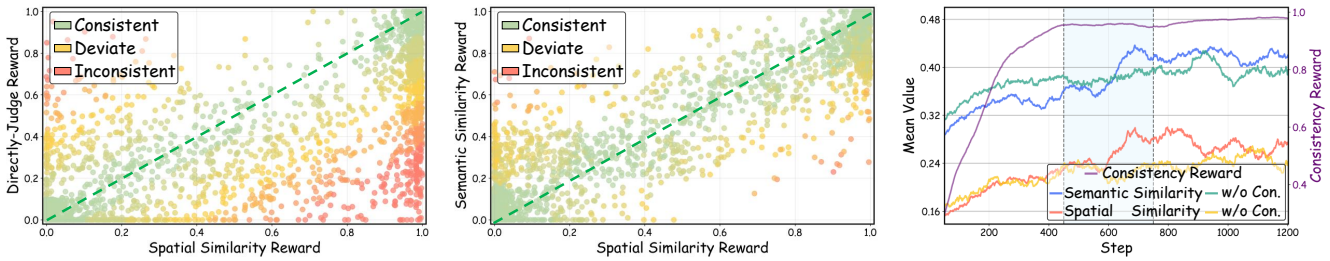
$$A_i = \frac{R_i - \text{mean}(R_j)}{\text{std}(R_j)}, \tag{1}$$

which quantifies how much this reply outperforms or underperforms its peers. During optimization, the same advantage A_i is broadcast to every token of reply i, and the policy network is updated with a PPO-like objective:

$$J_{\text{G}}(\theta) = \mathbb{E}\big[\min\big(r_{i,t}A_i,\ \text{clip}(r_{i,t}, 1-\varepsilon, 1+\varepsilon)A_i\big)\big], \tag{2}$$

where $r_{i,t}$ is the importance weight between the new and old policies, and the clipping threshold ε keeps the update conservative for stability.

Previous works [54, 89, 102] design reward functions that directly-judge whether texts are equal to enhance model's geolocation capabilities. However, this overlooks the characteristics of the geographic task because multiple locations can correspond to the same geographic position. Based on this insight, we develop the geo-similarity reward



(a) Correlation between Spatial Similarity Reward and Directly-Judge Reward

(b) Correlation between Spatial Similarity Reward and Semantic Similarity Reward

(c) Reward function value over training steps

Figure 4. **Discussion of reward functions.** The two scatter plots respectively reveal the reasons for the unreasonable directly-judge reward and the positive effect of semantic similarity reward. We also demonstrate the curve of reward value changes over training steps, reflecting the importance of applying consistency reward to enable the model to establish a complete reasoning framework.

that measures the similarity between the model's response and the correct answer. Specifically, it contains two components: spatial similarity R_{spa} and semantic similarity R_{sem}. Additionally, we also design the consistency reward R_{con} to maintain the integrity and consistency of CoT. Finally, the reward function can be expressed as:

$$R = a_1 R_{\text{spa}} + a_2 R_{\text{sem}} + a_3 R_{\text{con}}, \tag{3}$$

where a_1, a_2, a_3 are weights controlling the importance of three components, which are 1.5, 1.0 and 0.5, respectively.

4.2. Geo-Similarity Reward

Spatial similarity. To overcome the issue of non-unique mapping, we obtain the corresponding latitude $\hat{\lambda}$ and longitude $\hat{\phi}$ for the model's predictions through OpenCage [1] inverse geocoding. Then, the spatial distance $\mathcal{D}$ between the predicted location $(\hat{\lambda}, \hat{\phi})$ and the corresponding ground truth location (λ, ϕ) is calculated as below [39]:

$$\mathcal{D} = 2r \arcsin \sqrt{\sin^2 \frac{\Delta\phi}{2} + \cos\hat{\phi} \cos\phi \sin^2 \frac{\Delta\lambda}{2}}, \tag{4}$$

where $r = 6371\text{km}$ denotes the mean Earth radius and Δ is the difference. The spatial similarity reward is defined as:

$$R_{\text{spa}} = \exp\left(-\frac{\mathcal{D}\big((\hat{\lambda}, \hat{\varphi}), (\lambda, \varphi)\big)}{\tau}\right), \tag{5}$$

where τ is a temperature parameter controlling the reward sharpness. R_{spa} exhibits a nonlinear relationship with distance, specially increasing more slowly as the distance to the ground truth decreases. This encourages the model to identify a broad range of predictions before progressively narrowing it down, which aligns with human thinking.

Semantic similarity. To further enhance the model's ability to generate standardized address names and improve its robustness against ambiguous or alias geographic names, we introduce the semantic similarity reward [22, 65, 74]. We employ the multilingual semantic encoding model [74, 75] to encode each level i of the address into vector representations $\mathbf{h}_i^{pred}$ and $\mathbf{h}_i^{gt}$. Then, we compute the cosine similarity between each pair of embeddings:

$$s_i = \frac{\mathbf{h}_i^{pred} \cdot \mathbf{h}_i^{gt}}{\|\mathbf{h}_i^{pred}\| \, \|\mathbf{h}_i^{gt}\|}. \tag{6}$$

To suppress over low-quality results, we apply a threshold to filter similarity $\hat{s}_i$ greater than δ. Additionally, a hierarchical strategy is employed, which ensures that lower-level address components are only rewarded when higher-level predictions are not entirely wrong. Finally, the semantic similarity reward is computed as:

$$R_{\text{sem}} = \sum_{i=1}^{3} \alpha_i \, \hat{s}_i, \qquad \sum_i \alpha_i = 1, \tag{7}$$

where α_i represents the weight assigned to each level i. The semantic similarity reward effectively compensates for the limitations of spatial similarity in semantic understanding. It enables the model to achieve correct reward signals even in the presence of aliases, abbreviations, or translation variations. This encourages the generation of more standardized and semantically coherent geographic descriptions.

Discussion. In Fig. 4, we visualize the correlation between the directly-judge reward proposed by previous methods and our proposed semantic similarity reward and spatial similarity reward under the same dataset. Spatial similarity reward is a function directly related to distance difference, reflecting the model's geographic positioning capability. Therefore, the reward trend directly impacting the model's geolocation capabilities should align with it. However, Fig. 4 (a) clearly demonstrates that directly-judge reward exhibits severe inconsistencies with spatial similarity due to the non-unique mapping between descriptions and locations. Therefore, it is unreasonable to use it to enhance geolocation capabilities, and the results in the Tab. 3 also prove this. In contrast, as shown in Fig. 4 (b), the trends of semantic similarity rewards and spatial similarity rewards are highly consistent. This further explains the positive effect of semantic similarity rewards on the model.

Table 2. **Zero-shot Geolocation results on IM2GPS3K [34] and GeoSeek-Val.** ♢ indicates the model is not publicly accessible. ★, ♣ and ♠ represent LoRA [35], QLoRA [21] and full fine-tuning, respectively. Best and second-best results are in **bold** and underlined.

Models	Dataset	Size	IM2GPS3K (% @km)				GeoSeek-Val (% @km)				
			City 25km	Region 200km	Country 750km	Continent 2500km	City 25km	Region 200km	Country 750km	Continent 2500km	Geoscore 0-5k
GeoDecoder [19]♢	MP-16	4M	33.50	45.90	61.00	76.10	-	-	-	-	-
GeoCLIP [13]	MP-16	4M	34.47	50.65	69.97	83.82	11.82	29.04	56.13	79.13	3172.3
Translocator [72] ♢	MP-16	4M	31.10	46.70	58.90	80.10	-	-	-	-	-
PIGEOTTO [31] ♢	MP-16	4M	36.70	53.80	72.40	85.30	-	-	-	-	-
G3 [45] ♢	MP-16	4M	**40.94**	55.56	71.24	84.68	-	-	-	-	-
Qwen2.5-VL-7B [7]	-	-	21.93	29.93	43.08	60.68	3.57	7.95	20.03	44.96	1622.0
Qwen2.5-VL-32B [7]	-	-	23.36	30.35	43.46	59.70	4.06	7.43	16.10	39.06	1413.5
Gemma3-27B [83]	-	-	28.53	41.74	57.95	73.64	11.69	25.48	49.52	71.79	2704.0
InternVL3-14B [113]	-	-	20.10	27.51	40.89	59.01	3.43	7.67	18.69	43.50	1549.1
GeoReasoner★ [53]	GSV	133K	26.94	36.63	52.27	78.65	13.55	27.86	53.54	77.63	3083.9
GaGA [23] ♢ ♣	MG-Geo	5M	33.00	48.00	67.10	82.10	-	-	-	-	-
GRE-Suite-SFT [89] ♢ ♠	GRE-CoT	20K	29.30	44.78	62.43	78.81	-	-	-	-	-
GRE-Suite [89] ♢ ♠	GRE	30K	35.30	51.70	69.30	85.67	-	-	-	-	-
GLOBE [54] ♢ ♠	MP-16 Reason	33K	40.18	56.19	71.45	82.38	10.75	21.20	39.17	61.44	2412.5
GeoAgent-SFT (Ours)★	GeoSeek-CoT	10K	32.77	45.42	62.65	81.33	10.36	23.84	47.12	64.98	2968.9
GeoAgent (Ours) ★	GeoSeek	30K	40.75	**58.57**	**76.21**	**89.90**	**15.69**	**33.39**	**60.37**	**81.72**	**3314.1**

4.3. Consistency Reward

As shown in Fig. 3, one possible reason for inconsistent reasoning processes is that the geo-similarity reward represents the relationship between images and geographic locations. However, we also need to establish the relationship between images and geographic clues. This involves progressively narrowing down potential geographic areas before establishing a connection to a specific location. Therefore, we propose the consistency reward, assessed by a consistency agent that only obtains reasoning processes output by the GeoAgent but not its conclusions, as shown in Fig. 2. The consistency reward R_{con} is defined as:

$$R_{\text{con}} = \sum_i \mathbb{1}\left(\hat{y}_i = y_i\right) \cdot w_i \cdot p_i, \qquad \sum_i w_i = 1, \quad (8)$$

where $\mathbb{1}[\cdot]$ denotes the indicator function that evaluates to 1 if the model's predicted conclusion $\hat{y}_i$ matches the ground-truth conclusion y_i, and 0 otherwise. The weight w_i corresponds to the importance of the i-th geographic granularity level (i.e., country, region, or precise location). Additionally, p_i is a penalty term introduced to discourage the model from generating overly simplistic reasoning processes that could evade detection by the consistency agent (such as outputting only a final conclusion). This factor is a value proportional to the i-th level length ℓ_i of the reasoning process and can be expressed as:

$$p_i = \frac{1}{1 + \exp\left(-\lambda(\hat{\ell} - \mu)\right)}, \hat{\ell} = \frac{\ell_i - \min(\ell)}{\max(\ell) - \min(\ell)}, \quad (9)$$

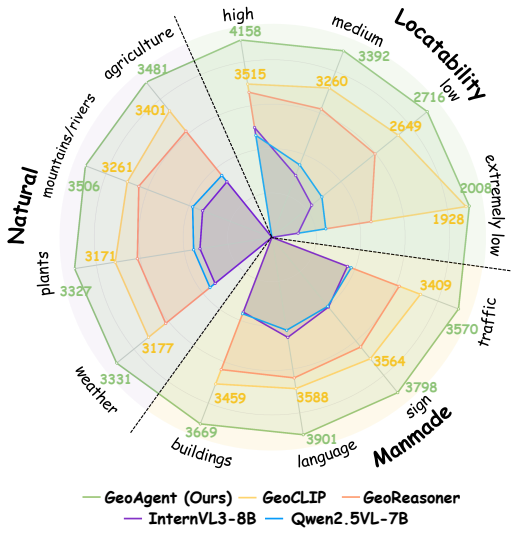


Figure 5. **GeoScore on GeoSeek-Val.** We compare different models in multiple locatabilities and geographic elements.

where λ and μ are the parameters of the control curve. The consistency reward models the relationships from images to geographic clues, from clues to layer-by-layer analyses, and finally to geographic locations. This ensures the integrity and consistency of the CoT at each level.

Discussion. As shown in Fig. 4, when training progresses, the consistency reward converges first. This enables the model to gradually establish a complete and consistent reasoning process. Compared to the model without consistency reward, both spatial and semantic similarity rewards remain lower than the existing model before the consistency reward converges. However, after convergence, these

Table 3. **Ablation study results reported on GeoSeek-Val.** Best and second-best results are in **bold** and underlined.

Models	SFT	COT	GRPO			Metrics		
			Spa. Reward	Sem. Reward	Con. Reward	City	Region	Country
Qwen2.5-VL-7B [7]						1.39	3.36	11.13
GeoAgent-SFT	✓					10.36	23.84	47.12
GeoAgent w/o Spa & Con	✓	✓		✓		11.99	26.56	55.55
GeoAgent w/o Sem & Con	✓	✓	✓			14.46	31.51	59.86
GeoAgent w/o Con	✓	✓	✓	✓		14.69	31.39	60.20
GeoAgent w/o Spa & Sem	✓	✓			✓	9.08	20.03	40.43
GeoAgent w/o SFT		✓	✓	✓	✓	13.39	23.94	58.23
GeoAgent	✓	✓	✓	✓	✓	**15.69**	**33.39**	**60.37**

Table 4. **Discussion of geo-similarity and directly-judge on GeoSeek-Val.** The judge means replacing the geo-similarity reward with directly-judge reward.

Models	Metrics		
	City	Region	Country
Qwen2.5-VL-7B [7]	1.39	3.36	11.13
GeoAgent-SFT	10.36	23.84	47.12
GeoAgent-SFT + Judge	12.35	26.87	50.81
GeoAgent-SFT + Judge + Con	11.04	26.13	51.76
GeoAgent w/o Con	14.69	31.39	60.20
GeoAgent	**15.36**	**32.53**	**60.37**

rewards begin to increase and surpass the model without consistency reward. This explains how consistency rewards positively impact the model's geolocation capabilities.

5. Experiments

5.1. Experimental details

Implementation details. GeoAgent is fine-tuned upon Qwen2.5-VL-7B [7] using LoRA [35] with rank of 64 and alpha of 128, while the consistency agent utilizes the GPTQ-INT4 [26] quantized version of Qwen3-32B [101].

Evaluation settings. An evaluation of GeoAgent is conducted using a public benchmark IMG2GPS3K [34] and GeoSeek-Val. In accordance with previous studies [54], we employ the City (25km), Region (200km), Country (750km) and Continent (2500km) accuracy on IM2GPS3K [34]. Furthermore, on GeoSeek-Val, we report the GeoScore, a widely utilized metric in the geolocation community. The scale ranges from 0 to 5000. The formula is as follows:

$$\text{GeoScore} = 5000 \times \exp\left(-\frac{10\,d}{d_{\max}}\right), \tag{10}$$

where d represents the distance between the predicted location and the actual location (in kilometers), while $d_{\max}$ denotes the scale. For global geolocation, $d_{\max}$ is typically set to 18,050 km. We utilize OpenCage [1] for geocoding and inverse geocoding.

5.2. Main Results

Performance on Public Benchmarks. We perform a comparative analysis of GeoAgent with multiple approaches in Tab. 2, including traditional methods, general VLLMs, and geolocation-specific VLLMs. With only using LoRA [35] fine-tuning (1.91% Paras), GeoAgent achieves a considerable performance improvement on the IM2GPS3K [111] (e.g., 72.40 → 76.21 on country accuracy), surpassing a series of full fine-tuning methods. It is also worth noting that GeoAgent demonstrates significantly greater performance improvements at the macro level compared to the fine level. This is because the geographical similarity reward gradually decelerates as distance decreases, prompting the model to first determine a coarse-grained location. In fact, in geolocation competition, country-level precision positioning is sufficient to outperform the most exceptional players.

Performance on GeoSeek-Val. We compare the accuracy of GeoAgent and other methods on GeoSeek-Val in Tab. 2. Also, we report their GeoScore on splits categorized by locatability and geographic elements in Fig. 5. The results demonstrate that GeoAgent significantly exceeds other models on the street-view geolocation benchmark (e.g., 56.13 → 60.37 on country accuracy). In addition, the results shown in Fig. 5 demonstrate that GeoAgent surpasses other models across all locatability and geographic element splits. This indicates that GeoAgent possesses a superior understanding of various geographic elements. Furthermore, the phenomenon where GeoAgent outperforms other models more significantly on splits with higher locatability suggests that GeoAgent's thinking patterns are closer to those of humans.

Dataset Quality. As shown in Tab. 2, our GeoAgent with only SFT achieves performance surpassing other methods with only SFT (GeoReasoner [53], GaGA [23] and GRE-suite-SFT [89]) despite operating on significantly less data (10K vs 20K, 133K, 5M). This fully demonstrates the high quality of the data annotated by GeoSeek-CoT's geographic experts and geolocation game players.

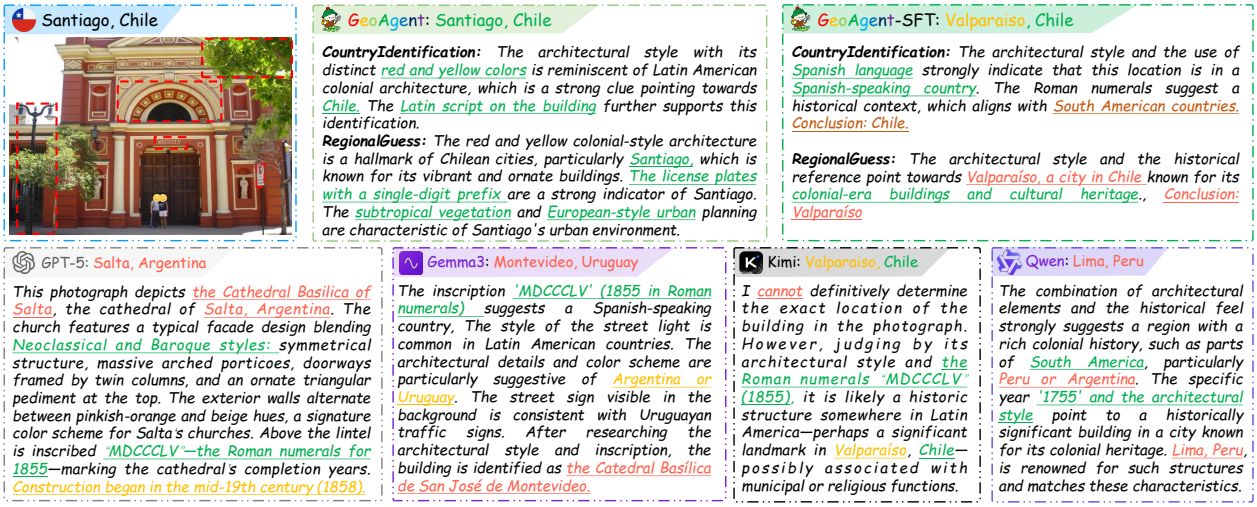


Figure 6. **Reasoning comparison of six different models (GPT-5 [69], Gemma3 [83], Kimi [84], Qwen2.5-VL-32B [7], GeoAgent and GeoAgent with only SFT).** Green: correct reasoning, Yellow: reasoning with certain issues, Red: reasoning that deviates significantly. Notably, brown highlights instances where the reasoning process is **incomplete** or **inconsistent** with the result.

5.3. Ablation Analysis

Ablation study of components. In Tab. 3, we discuss the effectiveness of the reward function we propose. It can be seen that R_{spa}, R_{sem}, and R_{con} each contribute to performance improvements. Among these, R_{spa} delivers a more significant performance improvement than R_{sem}. This is primarily because R_{spa} is a distance-dependent function that serves as a direct reward signal, whereas R_{sem} focuses on semantic robustness. When applied independently, R_{con} causes a slight decrease in performance. This can be attributed to it imposes constraints on consistency rather than directly addressing geolocation capabilities. However, when combined with R_{spa} and R_{sem}, it brings noticeable improvements, especially at the regional and city levels, where inconsistencies are more frequent than at the country level.

Cold Start. Using GeoSeek-CoT for SFT-based cold start enables the model to establish a reasoning framework, which aligns closely with human thinking patterns. The results in Tab. 3 indicate that cold starts can significantly enhance the base model's performance (e.g., 11.13 $\rightarrow$ 47.12 on country accuracy). At the same time, models lacking cold start capabilities will experience a decline in performance. This demonstrates the importance of establishing a reasoning framework during the cold start phase.

Discussion of Geo-Similarity and Directly-Judge. To demonstrate that the geo-similarity reward function better aligns with the characteristics of geographic tasks, we compare it with a reward function that directly-judges whether texts are identical at each level. As shown in Tab. 4, the performance improvement from the directly-judge reward is significantly smaller than that from the geo-similarity reward (e.g., 50.81 and 60.37 on country accuracy). This is because the directly-judge reward is overly strict, overlooking the model's efforts when it is not entirely correct but is moving toward the correct answer. Geo-similarity, on the other hand, evaluates the similarity between predictions and answers from both distance and semantic perspectives. It encourages the model's positive efforts while mitigating the negative impact of cross-language and aliasing issues on stability in geographic tasks. This underscores the necessity of designing training methods tailored to geographic characteristics for geographic tasks.

5.4. Qualitative Comparisons

As shown in Fig. 6, we compare the reasoning processes of GeoAgent, GeoAgent with only SFT, and other general VLLMs to the same image. Compared to other models, GeoAgent exhibits a clearer hierarchical reasoning process, enabling it to better capture the relationship between geographic characteristics and geolocation. Notably, after GRPO training, GeoAgent not only enhances its ability to identify geographic features but also overcomes the shortcomings of incomplete and inconsistent reasoning process.

6. Conclusions

This paper proposes GeoAgent, a model capable of thinking like humans and providing a geographic location. We propose GeoSeek, a novel dataset with annotated CoT from human geographic experts and geolocation game players. In training, we employ a two-stage training combining SFT with GRPO fine-tuning. Considering the non-uniqueness of descriptions and location mappings , we introduce the geo-similarity reward. In addition, we introduce the consistency reward to ensure the consistency of CoT. Experimental results demonstrate that GeoAgent achieves outstanding performance on multiple benchmarks.

7. Acknowledgments

This work was supported by NSFC (62522607, 62495061, and 62276145), the Fundamental Research Funds for the Central Universities (Nankai University), Shenzhen Science and Technology Program (JCYJ20240813114237048), "Science and Technology Yongjiang 2035" key technology breakthrough plan project (2025Z053), and Chinese government-guided local science and technology development fund projects (scientific and technological achievement transfer and transformation projects) (254Z0102G).

We sincerely thank Yue Zhang, H.M., Haowen He, Yuke Jun, and other experts in geography, as well as outstanding geolocation game players, for their valuable guidance, prompt design suggestions, and data support throughout the construction of the GeoSeek dataset.

We also thank Zhixiang Wang, Chilin Chen, Jincheng Shi, Liupeng Zhang, Yuan Gu, Yanghang Shao, Jinhua Zhang, Jiachen Zhu, Gucheng Qiuyue, Qingyang Guo, Jingchen Yang, Weilong Kong, Xinyuan Li, and Dawei Xu for their outstanding contributions in providing high-quality reasoning process data.

References

[1] OpenCage Geocoder. https://opencagedata.com/. Accessed: 2025-10-21. 5, 7

[2] Geoguessr - let's explore the world! https://www.geoguessr.com/. Accessed: 2025-10-14. 1, 3

[3] Tuxun:explore the world. https://tuxun.fun/. Accessed: 2025-10-14. 1, 3

[4] *Geolocation Estimation of Photos Using a Hierarchical Model and Scene Classification*, pages 575–592. Springer International Publishing, Cham, 2018. 1, 2

[5] Relja Arandjelović, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition, 2016. 2

[6] Guillaume Astruc, Nicolas Dufour, Ioannis Siglidis, Constantin Aronssohn, Nacim Bouia, Stephanie Fu, Romain Loiseau, Van Nguyen Nguyen, Charles Raude, Elliot Vincent, Lintao XU, Hongyu Zhou, and Loic Landrieu. Openstreetview-5m: The many roads to global visual geolocation, 2024. 2, 3, 4

[7] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 4, 6, 7, 8

[8] Gabriele Berton, Carlo Masone, and Barbara Caputo. Rethinking visual geo-localization for large-scale applications, 2022. 2

[9] Gabriele Moreno Berton, Valerio Paolicelli, Carlo Masone, and Barbara Caputo. Adaptive-attentive geolocalization from few queries: A hybrid approach. pages 2917–2926, 2021. 2

[10] Maksym Bondarenko, Rhorom Priyatikanto, Natalia Tejedor Garavito, Wenbin Zhang, Tom McKeen, Alexander Cunningham, Thea Woods, Jason Hilton, Duygu Cihan, Borys Nosatiuk, et al. The spatial distribution of population in 2015-2030 at a resolution of 30 arc (approximately 1km at the equator) r2025a version v1., 2025. 4

[11] Jun-Xiong Cai, Wensen Feng, Hao-Xiang Chen, and Tai-Jiang Mu. Sps: Accurate and real-time semantic positioning system based on low-cost dem maps. *IEEE Transactions on Image Processing*, 32:6401–6412, 2023. 1

[12] Ron Campos, Ashmal Vayani, Parth Parag Kulkarni, Rohit Gupta, Aizan Zafar, Aritra Dutta, and Mubarak Shah. GAEA: A Geolocation Aware Conversational Assistant, 2025. 3

[13] Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geolocalization, 2023. 2, 6

[14] Boyi Chen, Zhangyu Wang, Fabian Deuser, Johann Maximilian Zollner, and Martin Werner. Enhancing Contrastive Learning for Geolocalization by Discovering Hard Negatives on Semivariograms, 2025. 2

[15] Qiguang Chen, Libo Qin, Jinhao Liu, Dengyun Peng, Jiannan Guan, Peng Wang, Mengkang Hu, Yuhang Zhou, Te Gao, and Wanxiang Che. Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567*, 2025. 1

[16] Zhenfang Chen, Qinhong Zhou, Yikang Shen, Yining Hong, Zhiqing Sun, Dan Gutfreund, and Chuang Gan. Visual chain-of-thought prompting for knowledge-based visual reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1254–1262, 2024. 1

[17] Fenghua Cheng, Jinxiang Wang, Sen Wang, Zi Huang, and Xue Li. GeoGuess: Multimodal Reasoning based on Hierarchy of Visual Information in Street View, 2025. 2

[18] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017. 2

[19] Brandon Clark, Alec Kerrigan, Parth Parag Kulkarni, Vicente Vivanco Cepeda, and Mubarak Shah. Where we are and what we're looking at: Query based worldwide image geo-localization using hierarchies and scenes, 2023. 2, 6

[20] DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. 4

[21] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, 36: 10088–10115, 2023. 6

[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019. 5

[23] Zhiyang Dou, Zipeng Wang, Xumeng Han, Guorong Li, Zhipei Huang, and Zhenjun Han. Gaga: Towards interactive global geolocation assistant, 2025. 1, 2, 3, 6, 7

[24] Elias Dritsas and Maria Trigka. Remote sensing and geospatial analysis in the big data era: A survey. *Remote Sensing*, 17(3):550, 2025. 1

[25] Nicolas Dufour, David Picard, Vicky Kalogeiton, and Loic Landrieu. Around the world in 80 timesteps: A generative approach to global visual geolocation, 2024. 2

[26] Elias Frantar, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. Gptq: Accurate post-training quantization for generative pre-trained transformers. *arXiv preprint arXiv:2210.17323*, 2022. 7

[27] Yao Fu, Litu Ou, Mingyu Chen, Yuhao Wan, Hao Peng, and Tushar Khot. Chain-of-thought hub: A continuous effort to measure large language models' reasoning performance. *arXiv preprint arXiv:2305.17306*, 2023. 1

[28] Yixiao Ge, Haibo Wang, Feng Zhu, Rui Zhao, and Hongsheng Li. Self-supervising fine-grained region similarities for large-scale image localization, 2020. 2

[29] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. Pmlr, 2018. 1

[30] Lukas Haas, Silas Alberti, and Michal Skreta. Learning generalized zero-shot learners for open-domain image geolocalization, 2023. 2

[31] Lukas Haas, Michal Skreta, Silas Alberti, and Chelsea Finn. Pigeon: Predicting image geolocations. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12893–12902, Seattle, WA, USA, 2024. IEEE. 1, 2, 6

[32] Xiao Han, Chen Zhu, Xiangyu Zhao, and Hengshu Zhu. Swarm intelligence in geo-localization: A multi-agent large vision-language model collaborative framework, 2024. 1, 2, 3

[33] Stephen Hausler, Sourav Garg, Ming Xu, Michael Milford, and Tobias Fischer. Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition, 2021. 2

[34] James Hays and Alexei A. Efros. Im2gps: Estimating geographic information from a single image. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Anchorage, AK, USA, 2008. IEEE. 1, 2, 3, 6, 7

[35] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022. 6, 7

[36] Wenmiao Hu, Yichen Zhang, Yuxuan Liang, Xianjing Han, Yifang Yin, Hannes Kruppa, See-Kiong Ng, and Roger Zimmermann. Petalview: Fine-grained location and orientation extraction of street-view images via cross-view local search. In *Proceedings of the 31st ACM International Conference on Multimedia*, page 56–66. ACM, 2023. 2

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

[38] Sarah Ibrahimi, Nanne van Noord, Tim Alpherts, and Marcel Worring. Inside out visual place recognition, 2021. 2

[39] James Inman. *Navigation and nautical astronomy: For the use of British seamen*. F. and J. Rivington, 1849. 5

[40] International Road Federation (IRF). World Road Statistics Data Warehouse, 2025. Accessed: 2025-11-12. 4

[41] Mike Izbicki, Evangelos E Papalexakis, and Vassilis J Tsotras. *Exploiting the earth's spherical geometry to geolocate images*, pages 3–19. 2019. 1, 2

[42] Yuxiang Ji, Yong Wang, Ziyu Ma, Yiming Hu, Hailang Huang, Xuecai Hu, Guanhua Chen, Liaoni Wu, and Xiangxiang Chu. Thinking with Map: Reinforced Parallel Map-Augmented Agent for Geolocalization, 2026. 3

[43] Furong Jia, Lanxin Liu, Ce Hou, Fan Zhang, Xinyan Liu, and Yu Liu. Towards Interpretable Geo-localization: A Concept-Aware Global Image-GPS Alignment Framework, 2025. 2

[44] Furong Jia, Ling Dai, Wenjin Deng, Fan Zhang, Chen Hu, Daxin Jiang, and Yu Liu. SpotAgent: Grounding Visual Geo-localization in Large Vision-Language Models through Agentic Reasoning, 2026. 3

[45] Pengyue Jia, Yiding Liu, Xiaopeng Li, Yuhao Wang, Yantong Du, Xiao Han, Xuetao Wei, Shuaiqiang Wang, Dawei Yin, and Xiangyu Zhao. G3: An effective and adaptive framework for worldwide geolocalization using large multi-modality models, 2024. 1, 2, 3, 6

[46] Pengyue Jia, Seongheon Park, Song Gao, Xiangyu Zhao, and Yixuan Li. Georanker: Distance-aware ranking for worldwide image geolocalization, 2025. 1, 2, 3

[47] Hyo Jin Kim, Enrique Dunn, and Jan-Michael Frahm. Learned contextual feature reweighting for image geolocalization. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017. IEEE. 1, 2

[48] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213, 2022. 1

[49] Giorgos Kordopatis-Zilos, Panagiotis Galopoulos, Symeon Papadopoulos, and Ioannis Kompatsiaris. Leveraging efficientnet and contrastive learning for accurate global-scale location estimation, 2021. 1, 2

[50] Martha Larson, Mohammad Soleymani, Guillaume Gravier, Bogdan Ionescu, and Gareth JF Jones. The benchmarking initiative for multimedia evaluation: Mediaeval 2016. *IEEE MultiMedia*, 24(1):93–96, 2017. 2, 3

[51] Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, et al. Llava-onevision: Easy visual task transfer. *arXiv preprint arXiv:2408.03326*, 2024. 1

[52] Jiaao Li, Yixiang Huang, Ming Wu, Bin Zhang, Xu Ji, and Chuang Zhang. Clip-sp: Vision-language model with adaptive prompting for scene parsing. *Computational Visual Media*, 10(4):741–752, 2024. 1

[53] Ling Li, Yu Ye, Bingchuan Jiang, and Wei Zeng. Georeasoner: Geo-localization with reasoning in street views using a large vision-language model, 2024. 1, 3, 6, 7

[54] Ling Li, Yao Zhou, Yuxuan Liang, Fugee Tsung, and Jiaheng Wei. Recognition through reasoning: Reinforcing

image geo-localization with large vision-language models, 2025. 1, 2, 3, 4, 6, 7

[55] Qiujun Li, Zijin Xiao, Xulin Wang, Zhidan Ma, Cheng Yang, and Haifeng Li. LocationAgent: A Hierarchical Agent for Image Geolocation via Decoupling Strategy and Evidence from Parametric Knowledge, 2026. 3

[56] Xiang Li, Congcong Wen, Yuan Hu, and Nan Zhou. Rs-clip: Zero shot remote sensing scene classification via contrastive vision-language supervision. *International Journal of Applied Earth Observation and Geoinformation*, 124: 103497, 2023. 1

[57] Yuxi Li. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017. 1

[58] Yujie Li, Wenjia Xu, Guangzuo Li, Zijian Yu, Zhiwei Wei, Jiuniu Wang, and Mugen Peng. Unirs: Unifying multi-temporal remote sensing tasks through vision language models. *arXiv preprint arXiv:2412.20742*, 2024. 1

[59] Philipp Lindenberger, Paul-Edouard Sarlin, Jan Hosang, Matteo Balice, Marc Pollefeys, Simon Lynen, and Eduard Trulls. Scaling Image Geo-Localization to Continent Level, 2025. 2

[60] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023. 1

[61] Liu Liu and Hongdong Li. Lending orientation to neural networks for cross-view geo-localization, 2019. 2

[62] Liu Liu, Hongdong Li, and Yuchao Dai. Stochastic attraction-repulsion embedding for large scale image localization. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2570–2579, Seoul, Korea (South), 2019. IEEE. 2

[63] Mojgan Madadikhaljan and Michael Schmitt. Geolocation-aware deep coding: Infusing geolocation information into deep neural networks for remote sensing. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 93(1):3–18, 2025. 1

[64] Mohammad Sultan Mahmud, Joshua Zhexue Huang, Salman Salloum, Tamer Z Emara, and Kuanishbay Sadatdiynov. A survey of data partitioning and sampling methods to support big data analysis. *Big Data Mining and Analytics*, 3(2):85–101, 2020. 4

[65] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26, 2013. 5

[66] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015. 1

[67] OpenAI. Gpt-4o system card, 2024. 2

[68] OpenAI. Gpt-4.1, 2025. Accessed: 2025-05. 1

[69] OpenAI. Introducing GPT-5, 2025. Accessed: 2025-11-07. 8

[70] Chao Pang, Xingxing Weng, Jiang Wu, Jiayu Li, Yi Liu, Jiaxing Sun, Weijia Li, Shuai Wang, Litong Feng, Gui-Song Xia, et al. Vhm: Versatile and honest vision language model for remote sensing image analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6381–6388, 2025. 1

[71] Guohao Peng, Yufeng Yue, Jun Zhang, Zhenyu Wu, Xiaoyu Tang, and Danwei Wang. Semantic reinforced attention learning for visual place recognition, 2021. 2

[72] Shraman Pramanick, Ewa M. Nowara, Joshua Gleason, Carlos D. Castillo, and Rama Chellappa. Where in the world is this image? transformer-based geo-localization in the wild, 2022. 6

[73] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-tuning cnn image retrieval with no human annotation, 2018. 1, 2

[74] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019. 5

[75] Nils Reimers and Iryna Gurevych. paraphrase-multilingual-MiniLM-L12-v2: Multilingual Sentence Transformer Model, 2020. Accessed: 2025-11-12. 5

[76] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 1

[77] Paul Hongsuck Seo, Tobias Weyand, Jack Sim, and Bohyung Han. Cplanet: Enhancing image geolocalization by combinatorial partitioning of maps, 2018. 1, 2

[78] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. 3

[79] Zirui Song, Jingpu Yang, Yuan Huang, Jonathan Tonglet, Zeyu Zhang, Tao Cheng, Meng Fang, Iryna Gurevych, and Xiuying Chen. Geolocation with real human gameplay data: A large-scale dataset and human-like reasoning framework, 2025. 1, 2, 3

[80] Boyuan Sun, Jiaxing Zhao, Xihan Wei, and Qibin Hou. Llava-scissor: Token compression with semantic connected components for video llms. *arXiv preprint arXiv:2506.21862*, 2025. 1

[81] Hamed Taherdoost. Sampling methods in research methodology; how to choose a sampling technique for research. *International journal of academic research in management (IJARM)*, 5, 2016. 4

[82] Chameleon Team. Chameleon: Mixed-modal early-fusion foundation models. *arXiv preprint arXiv:2405.09818*, 2024. 1

[83] Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, et al. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*, 2025. 6, 8

[84] Kimi Team, Angang Du, Bohong Yin, Bowei Xing, Bowen Qu, Bowen Wang, Cheng Chen, Chenlin Zhang, Chenzhuang Du, Chu Wei, et al. Kimi-vl technical report. *arXiv preprint arXiv:2504.07491*, 2025. 8

[85] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth,

and Li-Jia Li. Yfcc100m: The new data in multimedia research. *Communications of the ACM*, 59(2):64–73, 2016. 3

[86] Akihiko Torii, Josef Sivic, Tomas Pajdla, and Masatoshi Okutomi. Visual place recognition with repetitive structures. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 883–890, 2013.

[87] Akihiko Torii, Relja Arandjelovic, Josef Sivic, Masatoshi Okutomi, and Tomas Pajdla. 24/7 place recognition by view synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1808–1817, 2015. 3

[88] Nam Vo, Nathan Jacobs, and James Hays. Revisiting im2gps in the deep learning era, 2017. 1, 2

[89] Chun Wang, Xiaoran Pan, Zihao Pan, Haofan Wang, and Yiren Song. Gre suite: Geo-localization inference via fine-tuned vision-language models and enhanced reasoning chains, 2025. 1, 2, 3, 4, 6, 7

[90] Qixuan Wang, Ning Li, Yiheng Chen, and Hainiu Zhu. Multitrans-lc: Multimodal fusion transformer for remote sensing land cover classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:133–138, 2025. 1

[91] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022. 1

[92] Frederik Warburg, Soren Hauberg, Manuel Lopez-Antequera, Pau Gargallo, Yubin Kuang, and Javier Civera. Mapillary street-level sequences: A dataset for lifelong place recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2623–2632, Seattle, WA, USA, 2020. IEEE. 3

[93] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022. 1

[94] Tobias Weyand, Ilya Kostrikov, and James Philbin. Planet - photo geolocation with convolutional neural networks. pages 37–55. 2016. 1, 2

[95] Scott Workman, Richard Souvenir, and Nathan Jacobs. Wide-area image geolocalization with aerial reference imagery, 2015. 1, 2

[96] Biao Wu, Meng Fang, Ling Chen, Ke Xu, Tao Cheng, and Jun Wang. Vision-Language Reasoning for Geolocalization: A Reinforcement Learning Approach, 2026. 3

[97] Rihui Xin, Han Liu, Zecheng Wang, Yupeng Zhang, Dianbo Sui, Xiaolin Hu, and Bingning Wang. Surrogate signals from format and length: Reinforcement learning for solving mathematical problems without ground truth answers. *arXiv preprint arXiv:2505.19439*, 2025. 2

[98] Chenhui Xu, Fuxun Yu, Michael J Bianco, Jacob Kovarskiy, Raphael Tang, Qi Zhang, Zirui Xu, Will LeVine, Brandon Dubbs, Heming Liao, et al. Geo-r1: Unlocking vlm geospatial reasoning with cross-view reinforcement learning. *arXiv preprint arXiv:2510.00072*, 2025. 1

[99] Shixiong Xu, Chenghao Zhang, Lubin Fan, Gaofeng Meng, Shiming Xiang, and Jieping Ye. Addressclip: Empowering vision-language models for city-wide image address localization, 2024. 2

[100] Shixiong Xu, Chenghao Zhang, Lubin Fan, Yuan Zhou, Bin Fan, Shiming Xiang, Gaofeng Meng, and Jieping Ye. Addressvlm: Cross-view alignment tuning for image address localization using large vision-language models, 2025. 2

[101] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025. 7

[102] Edward Yeo, Yuxuan Tong, Morry Niu, Graham Neubig, and Xiang Yue. Demystifying long chain-of-thought reasoning in llms. *arXiv preprint arXiv:2502.03373*, 2025. 4

[103] Sahiti Yerramilli, Nilay Pande, Rynaa Grover, and Jayant Sravan Tamarapalli. GeoChain: Multimodal Chain-of-Thought for Geographic Reasoning, 2025. 2

[104] Qiang Yi and Lianlei Shan. GeoLocSFT: Efficient Visual Geolocation via Supervised Fine-Tuning of Multimodal Foundation Models, 2025. 3

[105] Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Shiji Song, and Gao Huang. Does reinforcement learning really incentivize reasoning capacity in llms beyond the base model? *arXiv preprint arXiv:2504.13837*, 2025. 2

[106] Xiaohan Zhang, Xingyu Li, Waqas Sultani, Yi Zhou, and Safwan Wshah. Cross-view geo-localization via learning disentangled geometric layout correspondence, 2023. 2

[107] Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. Multimodal chain-of-thought reasoning in language models. *arXiv preprint arXiv:2302.00923*, 2023. 1

[108] Zilun Zhang, Tiancheng Zhao, Yulong Guo, and Jianwei Yin. Rs5m and georsclip: A large scale vision-language dataset and a large vision-language model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 1

[109] Zilun Zhang, Zian Guan, Tiancheng Zhao, Haozhan Shen, Tianyu Li, Yuxiang Cai, Zhonggen Su, Zhaojun Liu, Jianwei Yin, and Xiang Li. Geo-r1: Improving few-shot geospatial referring expression understanding with reinforcement fine-tuning. *arXiv preprint arXiv:2509.21976*, 2025. 1

[110] Jiaxing Zhao, Qize Yang, Yixing Peng, Detao Bai, Shimin Yao, Boyuan Sun, Xiang Chen, Shenghao Fu, Xihan Wei, Liefeng Bo, et al. Humanomni: A large vision-speech language model for human-centric video understanding. *arXiv preprint arXiv:2501.15111*, 2025. 1

[111] Zhongliang Zhou, Jielu Zhang, Zihan Guan, Mengxuan Hu, Ni Lao, Lan Mu, Sheng Li, and Gengchen Mai. Img2loc: Revisiting image geolocalization using multi-modality foundation models and image-based retrieval-augmented generation. pages 2749–2754, 2024. 2, 7

[112] Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*, 2023. 1

[113] Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, et al. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*, 2025. 6

2027届本科生公能素质评价加分材料

	类别	分数（自填）	审核结果
思想成长与身心发展	理想信念坚定	5	
	热心服务集体	0.8	
	先锋作用突出	0.5	
志愿服务与社会实践	志愿服务	1.75	
	社会实践	2	
艺体素质与技能特长	代表学院参与文体竞赛	0.25	
	文体竞赛获奖		
到国际组织挂职实习	---		
总计	---	10.3	

材料统计截止时间：

2021级 2024年6月30日 ☐
2022级 2025年6月29日 ☐
2023级 2026年7月05日 ☑

学号： 2312578

姓名： 金莫迪

所在班级： 2023物联网工程

（一）思想成长与身心发展（上限10分）

1、理想信念坚定（上限5分）

在校期间思想表现情况自述：

我在校期间思想表现坚定，作为南开大学密码与网络空间安全学院的一名中国共产党党员，，我始终牢记“允公允能，日新月异”的校训，将传承“豪密”红色基因与践行党员初心使命紧密结合。

在思想上，我坚持用党的创新理论武装头脑，深刻认识到网络安全是国家安全的重要基石。通过深入学习党的二十届三中全会精神，我进一步坚定了科技报国的理想信念，明确了将个人学术追求融入国家重大战略需求的责任担当。

在学习上，我刻苦钻研密码学与人工智能交叉领域的前沿技术，面对科研难题敢于迎难而上。同时，我的研究成果成功被国际计算机视觉顶级会议CVPR录用，这既是对我前期工作的肯定，更是鞭策。我努力在课题组中发挥先锋模范作用，与同学们分享科研心得，立志将核心技术掌握在自己手中。

在实践方面，我也积极投身社会实践与志愿服务，努力将所学转化为服务社会的实际行动。

同时，我也清醒认识到自身在理论联系实际方面仍有提升空间。未来，我将继续以党员标准严格要求自己，克服畏难情绪，练就过硬本领。

本人签字：

日 期 ：

辅导员评分：

辅导员签字：

日 期 ：

有无违反法律法规： 无

有无违反学校纪律处分规定： 无

有无发生思想品德问题并造成不良影响： 无

（一）思想成长与身心发展（上限10分）

2、热心服务集体

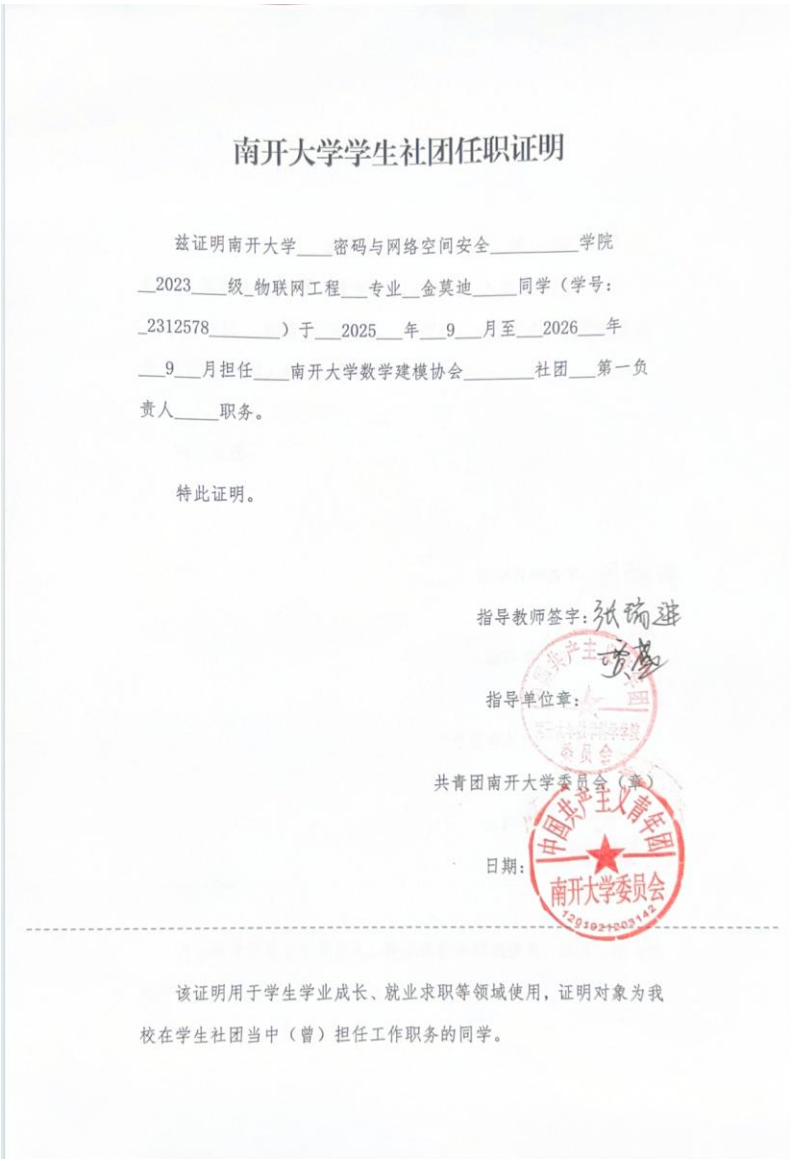
所属集体名称：密码与网络空间安全学院物联网工程2023级班

担任职务类别：其他班委

担任职务时间：2024-2025学年

（一）思想成长与身心发展（上限10分）

2、热心服务集体



南开大学学生社团任职证明

兹证明南开大学___密码与网络空间安全_____学院__2023___级_物联网工程___专业__金莫迪_____同学（学号：_2312578_____）于___2025___年___9___月至___2026___年___9___月担任____南开大学数学建模协会_____社团___第一负责人_____职务。

特此证明。

指导教师签字：张瑞进

指导单位章：

共青团南开大学委员会（章）

日期：

该证明用于学生学业成长、就业求职等领域使用，证明对象为我校在学生社团当中（曾）担任工作职务的同学。

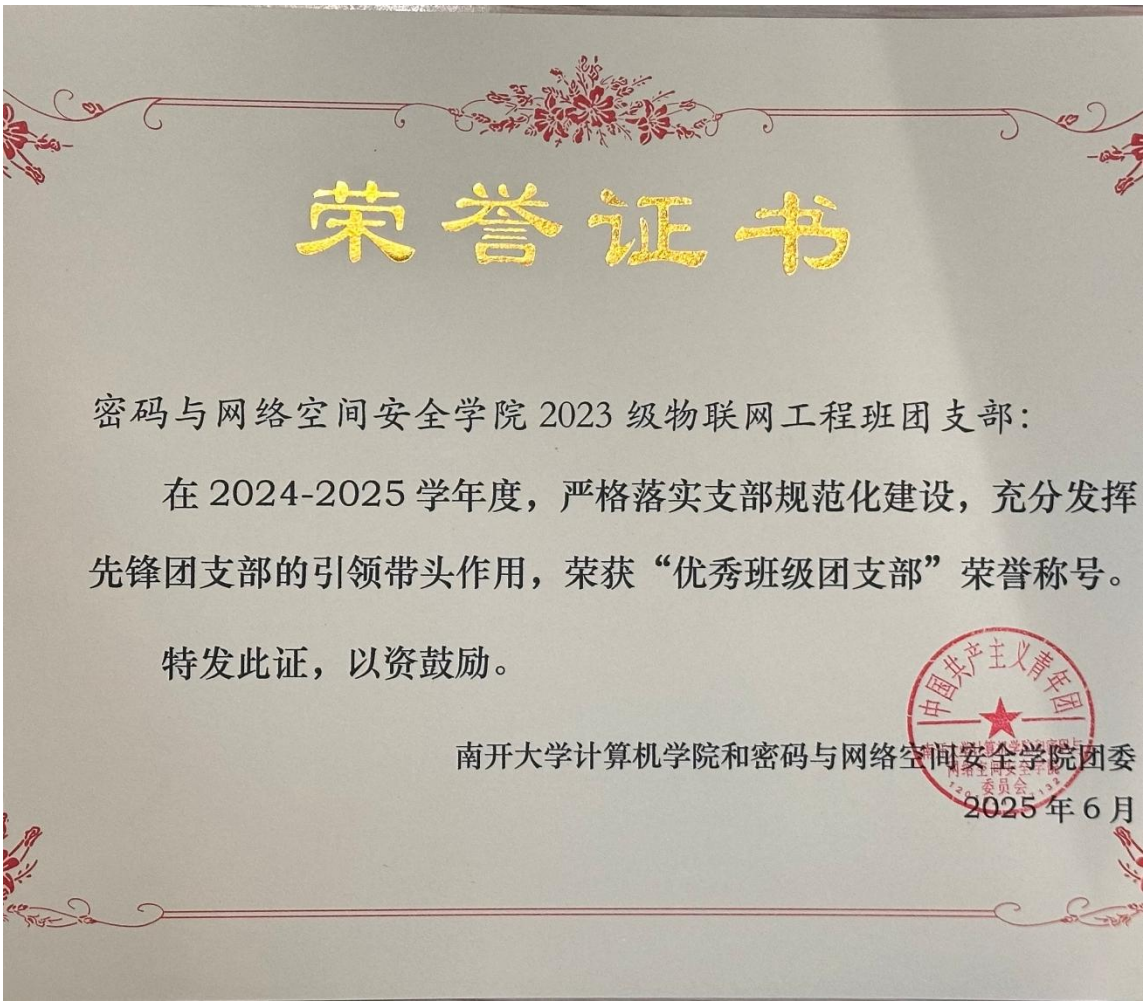
所属集体名称： 南开大学数学建模协会

担任职务类别： 其他社团社长

担任职务时间： 2025-2026学年

（一）思想成长与身心发展（上限10分）

3、先锋作用突出



荣誉证书

密码与网络空间安全学院2023级物联网工程班团支部：

在2024-2025学年度，严格落实支部规范化建设，充分发挥先锋团支部的引领带头作用，荣获“优秀班级团支部”荣誉称号。

特发此证，以资鼓励。

南开大学计算机学院和密码与网络空间安全学院团委

2025年6月

个人/团体奖项： 团体

荣誉所属学院： 密码与网络空间安全学院

荣誉名称： 优秀班级团支部

获奖时间： 2024-2025学年

荣誉所属等级： 院级

（一）思想成长与身心发展（上限10分）

3、先锋作用突出

计算机学院和密码与网络空间安全学院团委

关于 2025–2026 学年度团支部述职评议考核结果公示

经基础工作评价、团支部书记书面汇报，学院团委组织对各支部的日常工作情况、专项工作完成情况及系统平台数据情况进行全面考核评议，按学段、类别区分评选出下列支部为“优秀班团支部”。本学年获评过校级及以上荣誉的班团支部已自动授予院级荣誉（不占用本次评议考核院级荣誉名额），具体情况如下：

2022 级学博团支部
2024 级专博班团支部
2024 级学硕班团支部
2025 级专硕 1 班团支部
计算机学院 2023 级计算机科学与技术 1 班团支部
计算机学院 2023 级计算机科学与技术 3 班团支部
密码与网络空间安全学院 2023 级物联网工程班团支部
密码与网络空间安全学院 2023 级信息安全班团支部
计算机学院 2024 级计算机卓越班团支部
密码与网络空间安全学院 2024 级信息安全、法学双学位班团支部
密码与网络空间安全学院 2024 级密码科学与技术班团支部
计算机学院工科试验班模拟 1-1 班团支部
密码与网络空间安全学院豪密特色班团支部
密码与网络空间安全学院信息安全法学双学位班团支部

公示时间为 2026 年 7 月 1 日-7 月 3 日。对以上结果如有异议，请以书面形式联系学院 623 团委办公室。

南开大学计算机学院和密码与网络空间安全学院团委

2026 年 7 月 1 日

个人/团体奖项： 团体

荣誉所属学院： 密码与网络空间安全学院

荣誉名称： 优秀团支部

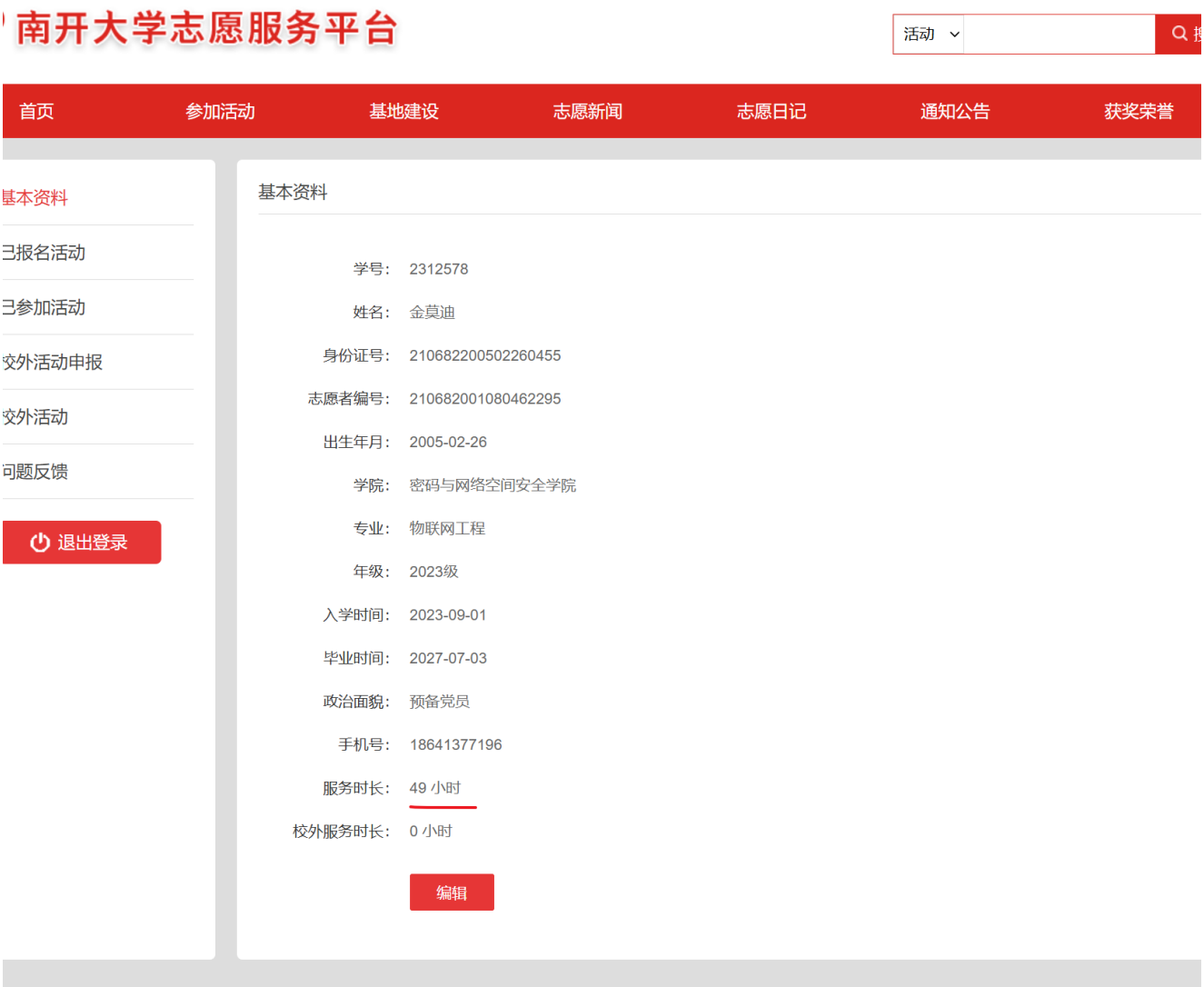
获奖时间： 2025-2026学年

荣誉所属等级： 院级

（二）志愿服务与社会实践（上限6分）

1、志愿服务

（1）志愿服务时长证明



前三学年内志愿服务总时长（小时）： 49

除公能实践课要求外志愿服务总时长（小时）： 35.5

（二）志愿服务与社会实践（上限6分）

1、志愿服务

（1）除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）

-数学文化节趣味游戏志愿活动-
志愿服务证明

金莫迪 同学：

于2026年03月14日08:00至2026年03月14日13:30，在化节趣味游戏志愿活动”活动中从事志愿服务工作，共计4小工作期间认真负责，发扬了奉献友爱互助进步的志愿者精神良好的南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2026

第36届“校长杯”羽毛球赛（津南校区）学生裁判-
志愿服务证明

金莫迪 同学：

于2023年11月01日08:00至2023年12月09日08:00，在学第36届“校长杯”羽毛球赛（津南校区）学生裁判”活动中服务工作，共计8.5小时。工作期间认真负责，发扬了奉献进步的志愿者精神，树立了良好的南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2023年

3“西部温暖”计划志愿活动线上培训
志愿服务证明

金莫迪 同学：

于2023年11月03日16:00至2023年11月03日18:00，1.3“西部温暖”计划志愿活动线上培训会”活动中从事志愿服共计0.5小时。工作期间认真负责，发扬了奉献友爱互助愿者精神，树立了良好的南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2023年

志愿服务时长（小时）：	志愿服务时长（小时）：	志愿服务时长（小时）：
4	8.5	0.5

（二）志愿服务与社会实践（上限6分）

1、志愿服务

（1）除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）

1.4津南校区“西部温暖”计划志愿活

志愿服务证明

金莫迪 同学：

于2023年11月04日08:00至2023年11月04日18:00，1.4津南校区“西部温暖”计划志愿活动”活动中从事志愿服共计3小时。工作期间认真负责，发扬了奉献友爱互助进步精神，树立了良好的南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2023年

-图书馆志愿讲解员招募-

志愿服务证明

金莫迪 同学：

于2024年04月19日08:00至2024年05月16日18:00，志愿讲解员招募”活动中从事志愿服务工作，共计1.5小时。同认真负责，发扬了奉献友爱互助进步的志愿者精神，南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2024

-南开书屋-

志愿服务证明

金莫迪 同学：

于2024年04月28日08:00至2024年04月30日20:00，在屋”活动中从事志愿服务工作，共计3小时。工作期间认真负发扬了奉献友爱互助进步的志愿者精神，树立了良好的南开者形象。

特此证明。

南开大学社会实践和志愿服务督导

2024

志愿服务时长（小时）：

3

志愿服务时长（小时）：

1.5

志愿服务时长（小时）：

3

（二）志愿服务与社会实践（上限6分）

1、志愿服务

（1）除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）

-第十一届公益晨跑-

志愿服务证明

金莫迪 同学：

于2024年05月07日08:00至2024年05月19日08:00，在于届公益晨跑”活动中从事志愿服务工作，共计8.5小时。工真负责，发扬了奉献友爱互助进步的志愿者精神，树立了大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2024

-新阳光病房支教（3月）-

志愿服务证明

金莫迪 同学：

于2025年02月18日08:00至2025年03月18日08:00，在病房支教（3月）”活动中从事志愿服务工作，共计12小时间认真负责，发扬了奉献友爱互助进步的志愿者精神，树南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2025

-数学建模大家谈系列活动-

志愿服务证明

金莫迪 同学：

于2025年11月15日10:00至2025年12月27日16:00，在模大家谈系列活动”活动中从事志愿服务工作，共计5小时。间认真负责，发扬了奉献友爱互助进步的志愿者精神，树南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导

2025年

志愿服务时长（小时）：

8.5

志愿服务时长（小时）：

12

志愿服务时长（小时）：

5

（二）志愿服务与社会实践（上限6分）

1、志愿服务

（1）除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）

-数学建模大家谈系列活动-

志愿服务证明

金莫迪 同学：

于2025年12月06日14:00至2025年12月13日17:00，在
模大家谈系列活动”活动中从事志愿服务工作，共计3小时。
间认真负责，发扬了奉献友爱互助进步的志愿者精神，树立
南开大学志愿者形象。

特此证明。

南开大学社会实践和志愿服务督导
2025年

志愿服务时长（小时）：

3

志愿服务时长（小时）：

志愿服务时长（小时）：

（二）志愿服务与社会实践（上限6分）

2、社会实践（上限为2分）

（1）社会实践次数证明

大一寒假是否参加社会实践：否

大二寒假是否参加社会实践：是

大三寒假是否参加社会实践：是

大一暑假是否参加社会实践：是

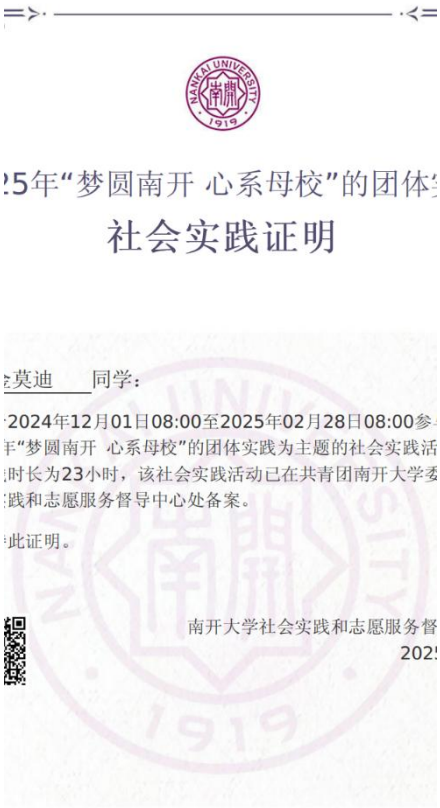
大二暑假是否参加社会实践：是

大三暑假是否参加社会实践：否

（二）志愿服务与社会实践（上限6分）

2、社会实践（上限为2分）

（1）社会实践证明



25年“梦圆南开 心系母校”的团体
社会实践证明

莫迪 同学：

2024年12月01日08:00至2025年02月28日08:00参
年“梦圆南开 心系母校”的团体实践为主题的社会实践活
时长为23小时，该社会实践活动已在共青团南开大学委
践和志愿服务督导中心处备案。

此证明。

南开大学社会实践和志愿服务督
202

-2026年寒假母校回访专题实践-
社会实践证明

莫迪 同学：

2026年01月01日08:00至2026年03月31日08:00参
年寒假母校回访专题实践为主题的社会实践活动，社会
小时，该社会实践活动已在共青团南开大学委员会社会
服务督导中心处备案。

此证明。

南开大学社会实践和志愿服务督
202

-2024年高考招生咨询志愿者-
社会实践证明

莫迪 同学：

2024年06月15日08:00至2024年07月15日08:00参
年高考招生咨询志愿者为主题的社会实践活动，社会实
时，该社会实践活动已在共青团南开大学委员会社会实
督导中心处备案。

此证明。

南开大学社会实践和志愿服务督
202

（二）志愿服务与社会实践（上限6分）

2、社会实践（上限为2分）

（1）社会实践证明

负责人信息	姓名：严景云（老师） 性别：女 学院：教务部 联系方式：13672072191 邮箱：nkyanjy@nankai.edu.cn
活动类型	团体活动
实践类型	日常社会实践专项
指导单位	教务处
拟实践项目	2025年高考招生咨询志愿者
拟实践地点	全国
活动时间	2025-06-15 00:00到2025-07-15 00:00
拟实践时长	100 小时
是否新建/维护南开书屋	否
是否新建/运维社会实践基地	否
是否依托"师生四同"课题	否
是否申请参与"莲子计划"	否
是否为服务学习课程实践项目	否
是否为2023年"校长杯"创新创业大赛参赛项目	否
是否为在研本科生创新科研计划项目（"国创百项"）	否
是否依托社团开展活动	否
实践选题意义	高考后和报志愿阶段的招生咨询逐渐转变为"以线下咨询为主，线上宣传为辅"的新常态,竞争形势与工作难度日益加剧，招生咨询人员需求进一步加大。招生咨询人员代表我校形象，其工作水平、能力和态度，很大程度会直接影响考生的志愿选择。招生办公室特组织有意愿参与高考招生的在校生在考后返回各自家乡、中学参与招生宣传，服务高考考生与家长，进行高考志愿填报指导，以吸引更多优秀考生选择南开。
前期准备方案	高考后和报志愿阶段的招生咨询逐渐转变为“以线下咨询为主，线上宣传为辅”的新常态,竞争形势与工作难度日益加剧，招生咨询人员需求进一步加大。招生咨询人员代表我校形象，其工作水平、能力和态度，很大程度会直接影响考生的志愿选择。招生办公室特组织有意愿参与高考招生的在校生在考后返回各自家乡、中学参与招生宣传，服务高考考生与家长，进行高考志愿填报指导，以吸引更多优秀考生选择南开。
完成项目的条件与保证	高考后和报志愿阶段的招生咨询逐渐转变为“以线下咨询为主，线上宣传为辅”的新常态,竞争形势与工作难度日益加剧，招生咨询人员需求进一步加大。招生咨询人员代表我校形象，其工作水平、能力和态度，很大程度会直接影响考生的志愿选择。招生办公室特组织有意愿参与高考招生的在校生在考后返回各自家乡、中学参与招生宣传，服务高考考生与家长，进行高考志愿填报指导，以吸引更多优秀考生选择南开。
实践进度安排	高考后和报志愿阶段的招生咨询逐渐转变为“以线下咨询为主，线上宣传为辅”的新常态,竞争形势与工作难度日益加剧，招生咨询人员需求进一步加大。招生咨询人员代表我校形象，其工作水平、能力和态度，很大程度会直接影响考生的志愿选择。招生办公室特组织有意愿参与高考招生的在校生在考后返回各自家乡、中学参与招生宣传，服务高考考生与家长，进行高考志愿填报指导，以吸引更多优秀考生选择南开。
实践预期成果	高考后和报志愿阶段的招生咨询逐渐转变为“以线下咨询为主，线上宣传为辅”的新常态,竞争形势与工作难度日益加剧，招生咨询人员需求进一步加大。招生咨询人员代表我校形象，其工作水平、能力和态度，很大程度会直接影响考生的志愿选择。招生办公室特组织有意愿参与高考招生的在校生在考后返回各自家乡、中学参与招生宣传，服务高考考生与家长，进行高考志愿填报指导，以吸引更多优秀考生选择南开。
经费预算	

类别	数量	单价	总金额
交通			
住宿			
材料	\	\	
其他	\	\	
总计	\	\	0.00

认定时长	32 小时

（二）志愿服务与社会实践（上限6分）

2、社会实践（上限为2分）

（2）社会实践获荣誉称号

为推动我校招生工作开展，鼓励更多学生参与到招生宣传的社会实践当中来，**南开大学招生办公室**牵头申请，**南开大学招生志愿者协会**组织开展了**“南风追梦”母校回访寒假社会实践活动。** 活动过程中，同学们精心准备宣讲材料，结合自身在南开的学习成长经历，通过线上线下多种渠道，向高中学生及家长系统介绍南开大学的办学特色、学科优势、培养体系及招生政策，有效搭建起南开与各地中学之间的沟通桥梁，提升了我校招生宣传工作的精细化和丰富性。

经过线上线下覆盖情况、宣讲内容丰富程度、宣讲后反馈情况等方面的综合考量，决定对本年度寒假社会实践先进队伍进行表彰。表彰名单如下：

□ 一等奖队伍

辽宁省实验中学
大连市第二十四中学
西安高新第一中学
天津市武清区杨村第一中学

□ 二等奖队伍

天津市宝坻区第一中学
忻州市第一中学校
武汉市洪山高级中学
兰州市三十三中学
福建省南平第一中学
鄂尔多斯市第一中学

□ 三等奖队伍

社会实践活动名称：南风追梦母校回访寒假社会实践活动

获奖时间：2026年6月

获奖名称：三等奖

是否为第一负责人：否

获奖等级：校级

（二）志愿服务与社会实践（上限6分）

2、社会实践（上限为2分）

（2）社会实践获荣誉称号

固原市第一中学
安徽省临泉县第一中学
铜仁市第一中学
西南大学附属中学校
石家庄精英中学
郑州外国语学校
西安高新第一中学
天津市静海区第一中学
四川省绵阳中学
太原市外国语学校
海南中学
辽宁省实验中学

2026年高考来临之际，南开大学高考集中咨询季招生咨询工作已全面启动。诚挚期待更多心怀热忱的南开学子主动加入暑期高考咨询志愿者队伍，以亲身经历分享求学故事，以朋辈力量吸引优秀学子逐梦南开，为学校招生工作高质量发展注入青春力量！

随招生组回家乡学生报名

在校电话咨询学生报名

社会实践活动名称：南风追梦母校回访寒假社会实践活动

获奖时间：2025年9月

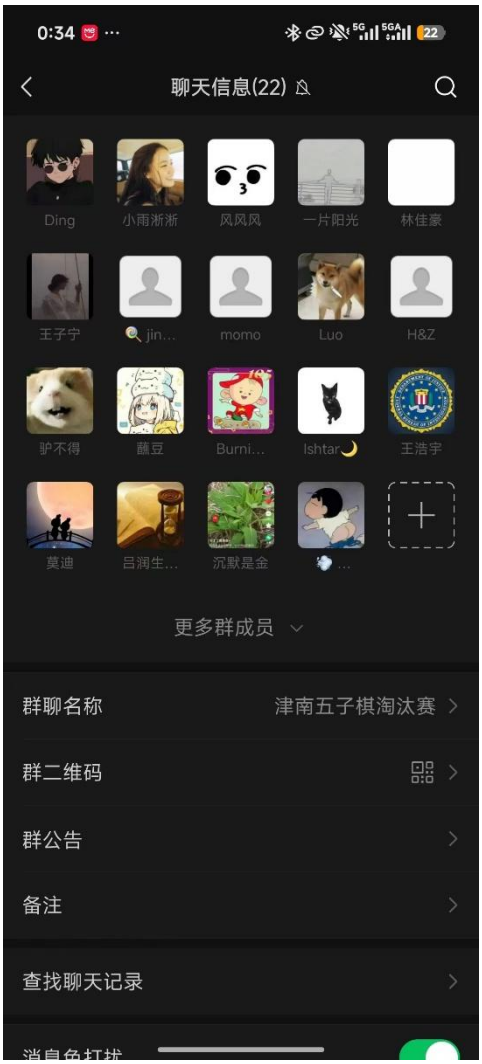
获奖名称：一等奖

是否为第一负责人：否

获奖等级：校级

（三）艺体素质与技能特长（上限2分）

1、代表学校、学院参加艺术体育活动



参加活动名称：五子棋校长杯

参加时间：2025年3月

所属学院：密码与网络空间安全学院

证明人：