NeSyGeo: A Neuro-Symbolic Framework for Multimodal Geometric Reasoning Data Generation

Weiming Wu [1] **Zi-Kang Wang** [1] **Jin Ye** [1] **Zhi Zhou** [2] **Yu-Feng Li** [2 3] **Lan-Zhe Guo** [1 3]

Abstract

Obtaining large-scale, high-quality data with reasoning paths is crucial for improving the geometric reasoning capabilities of multi-modal large language models (MLLMs). However, existing data generation methods inevitably face diversity and numerical generalization limitations. To address these limitations, we propose NeSyGeo, a novel neuro-symbolic framework for generating geometric reasoning data. First, we propose a domain-specific language grounded in the entity–relation–constraint paradigm to comprehensively represent all components of plane geometry, along with generative actions defined within this symbolic space. We then design a symbolic–visual–text pipeline that synthesizes symbolic sequences, maps them to corresponding visual and textual representations, and generates diverse question–answer (Q&A) pairs using large language models (LLMs). Based on this framework, we construct NeSyGeo-CoT and NeSyGeo-Caption datasets, containing 100k samples, and release a new benchmark NeSyGeo-Test for evaluating geometric reasoning abilities in MLLMs. Experiments demonstrate that the proposal significantly and consistently improves the performance of multiple MLLMs under both reinforcement and supervised fine-tuning. With only 4k samples and two epochs of reinforcement fine-tuning, base models achieve improvements of up to +15.8% on MathVision, +8.4% on MathVerse, and +7.3% on GeoQA. Notably, a 4B model can be improved to outperform an 8B model from the same series on geometric reasoning tasks.

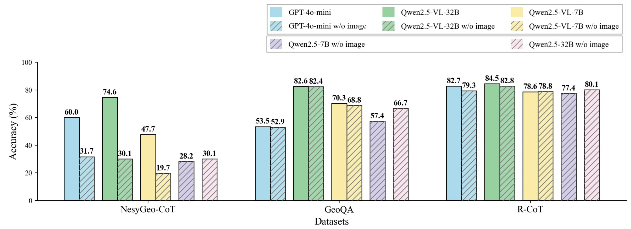


Figure 1. Performance comparison of different MLLMs and LLMs with and without image input in several geometry datasets. The minimal or negligible drops observed upon image removal in GeoQA and R-CoT raise concerns regarding the utilization of visual information for geometric reasoning.

1. Introduction

Improving the visual reasoning capabilities of MLLMs has garnered significant attention recently (Liu et al., 2023; Alayrac et al., 2022; Achiam et al., 2023; Li et al., 2021; Jiang et al., 2024; Zhang et al., 2025a; Wang et al., 2024a; Wu et al., 2024; Liang et al., 2024), with models like InternVL (Chen et al., 2024) and the QwenVL series (Wang et al., 2024c; Bai et al., 2025) demonstrating significant enhancements in visual-semantic comprehension through their multimodal capabilities. Among various visual reasoning tasks, geometric mathematical reasoning is crucial for evaluating the reasoning performance of MLLMs (Yan et al., 2025; 2024), as it requires a deep integration of spatial perception, symbolic understanding, and logical deduction. To enhance such reasoning abilities, existing approaches (Zhang et al., 2024d; 2025b; Zhao et al., 2025) primarily rely on fine-tuning base models using reinforcement learning (RL) or supervised fine-tuning (SFT) on specialized geometric reasoning datasets. These methods depend heavily on the availability of large-scale, high-quality geometric reasoning data, which is often costly and time-consuming to construct manually. Therefore, automatic data generation for geometric reasoning has emerged as a promising and actively explored direction, aiming to alleviate data scarcity and further improve the reasoning abilities of MLLMs.

Existing approaches for generating datasets in geometric tasks can be broadly classified into four categories. **Text**

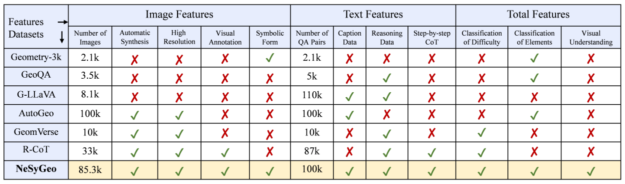
[1] School of Intelligence Science and Technology, Nanjing University, Nanjing, China [2] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China [3] School of Artificial Intelligence, Nanjing University, Nanjing, China. Correspondence to: Lan-Zhe Guo <guolz@lamda.nju.edu.cn>.

The second AI for MATH Workshop at the 42nd International Conference on Machine Learning, Vancouver, Canada.

augmentation methods like G-LLaVA (Gao et al., 2025) primarily mutate the conditions of existing datasets through equivalent condition transformation and numerical scaling. However, this approach fails to address the scalability of image generation. **Template-based methods** (Deng et al., 2024; Zhang et al., 2024d; Kazemi et al., 2024), use predefined geometric templates with fixed topologies, simplifying synthesis but constraining diversity by reducing the geometric space to limited combinations. **Solver-based methods** (Huang et al., 2025; Zhang et al., 2024b) inspired by symbolic prover AlphaGeometry (Trinh et al., 2024), leverage formal languages for synthesis but lack metric details (e.g., angles, lengths, areas), restricting multimodal data to descriptive annotations and limiting numerical reasoning applications. **Tool-based methods** attempt to generate codes for tools like GeoGebra or MATLAB via LLMs. However, even advanced models struggle to ensure correctness with ambiguous natural language instructions and complex geometric spaces. In summary, existing methods grapple with issues of image scalability, limited geometric diversity, a lack of precise numerical information, and challenges in ensuring the reliability of generated content.

Beyond the challenges in data synthesis methodologies, current geometric reasoning datasets present several limitations that impede the advancement of MLLMs. A primary limitation stems from the often inadequate quality and low resolution of the provided images. Such inputs frequently fall below the optimal requirements of visual encoders (Radford et al., 2021; Lin et al., 2025; Liu et al., 2023), hindering the extraction of crucial fine-grained visual features and discriminative information essential for robust multimodal reasoning. Furthermore, our analysis reveals notable information redundancy between the textual and visual modalities in many current datasets. As shown in Figure 1, our comparative experiments demonstrate minimal or negligible accuracy drops upon image removal. This finding emphasizes the urgent need for a dataset that effectively separates textual and visual information and provides high-quality images to promote MLLMs' visual perception and logical reasoning performance.

To address these challenges, we propose **NeSyGeo**, a neuro-symbolic framework for synthesising high-quality multimodal geometric reasoning datasets. NeSyGeo integrates three components: 1) A formal geometric symbolic space defined by a domain-specific language (DSL), capturing primitive entities (points, lines, circles), topological relations (parallelism, incidence, perpendicularity), and metric constraints (angles, lengths), enabling diverse geometric configurations via systematic sampling within constrained parametric bounds. 2) A bidirectional conversion engine that transforms symbolic constructs into decoupled modalities, producing annotated vector graphics paired with concise textual axioms. 3) A causal Q&A pairs and theorem-grounded Chain-of-Thought (CoT) sequences generator that effectively merges neural reasoning with symbolic verification. To our knowledge, we are the first to develop a neuro-symbolic framework for producing multimodal reasoning data.

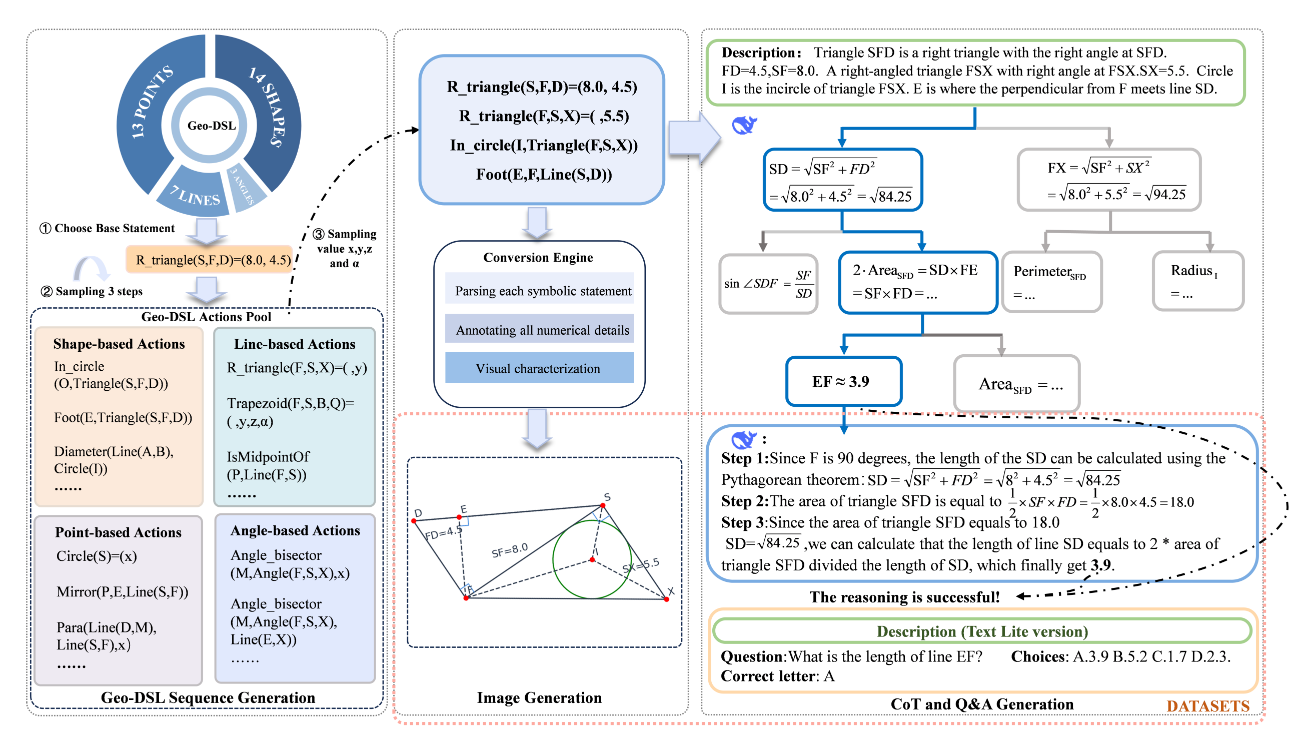


Features → Datasets ↓	Image Features					Text Features				Total Features		
	Number of Images	Automatic Synthesis	High Resolution	Visual Annotation	Symbolic Form	Number of QA Pairs	Caption Data	Reasoning Data	Step-by-step CoT	Classification of Difficulty	Classification of Elements	Visual Understanding
Geometry-3k	2.1k	✗	✗	✗	✓	2.1k	✗	✗	✗	✗	✓	✗
GeoQA	3.5k	✗	✗	✗	✗	5k	✗	✓	✗	✗	✓	✗
G-LLaVA	8.1k	✗	✗	✗	✗	110k	✓	✓	✗	✗	✗	✗
AutoGeo	100k	✓	✓	✗	✗	100k	✓	✗	✗	✗	✓	✗
GeomVerse	10k	✓	✓	✗	✗	10k	✗	✓	✗	✓	✗	✗
R-CoT	33k	✓	✓	✓	✗	87k	✗	✓	✓	✓	✗	✗
NeSyGeo	85.3k	✓	✓	✓	✓	100k	✓	✓	✓	✓	✓	✓

Figure 2. Comparison of dataset characteristics synthesized by our method and other popular synthesis approaches. "High Resolution" denotes average image pixels exceeding 336×336. "Symbolic Form" refers to the symbolic meta-information associated with the image. "Classification of Elements" signifies categorization by geometric elements. "Visual Understanding" represents the mitigation of image-text redundancy for stronger visual grounding in reasoning. More specific examples of different methods are in Appendix A.

Our framework enhances the diversity and validity of generated geometric reasoning data while effectively mitigating information redundancy and the underutilization of visual signals during training. Specifically, the comprehensive Geo-DSL and its expansive symbolic synthesis action-space promote diverse and well-grounded image generation. Meanwhile, our CoT sequence generator, powered by LLMs' strong reasoning and language capabilities, conducts a backwards search across the geometric space to construct Q&A pairs, thereby enriching textual diversity. The unique identification of geometric elements via our symbolic language and dedicated conversion engine ensures visual validity. In parallel, a bidirectional cross-validation process using expert LLMs ensures textual validity. By strategically distributing complementary information across image and text modalities, our approach encourages MLLMs to actively engage with visual information when solving problems, enhancing their abilities to perceive visually and effectively utilize images.

Leveraging the NeSyGeo pipeline, we construct two training datasets, NeSyGeo-Caption and NeSyGeo-CoT, comprising 100k samples. NeSyGeo-Caption aims to improve the perceptual understanding of geometric elements, while NeSyGeo-CoT primarily focuses on enhancing logical reasoning. The key characteristics of our dataset compared to other popular multimodal geometric datasets are presented in Figure 1. Additionally, we develop an evaluation set, NeSyGeo-Test, with 2668 Q&A pairs, enabling a thorough assessment of the geometric reasoning capabilities of mainstream MLLMs. We conducted an extensive and compre-


13 POINTS
14 SHAPES
7 LINES
3 ANGLES
Geo-DSL
① Choose Base Statement
R_triangle(S,F,D)=(8.0, 4.5)
② Sampling 3 steps
③ Sampling value x,y,z and α
Geo-DSL Actions Pool
Shape-based Actions
In_circle (O,Triangle(S,F,D))
Foot(E,Triangle(S,F,D))
Diameter(Line(A,B), Circle(I))
……
Line-based Actions
R_triangle(F,S,X)=(,y)
Trapezoid(F,S,B,Q)= (,y,z,α)
IsMidpointOf (P,Line(F,S))
……
Point-based Actions
Circle(S)=(x)
Mirror(P,E,Line(S,F))
Para(Line(D,M), Line(S,F),x)
……
Angle-based Actions
Angle_bisector (M,Angle(F,S,X),x)
Angle_bisector (M,Angle(F,S,X), Line(E,X))
……
Geo-DSL Sequence Generation
R_triangle(S,F,D)=(8.0, 4.5)
R_triangle(F,S,X)=(,5.5)
In_circle(I,Triangle(F,S,X))
Foot(E,F,Line(S,D))
Conversion Engine
Parsing each symbolic statement
Annotating all numerical details
Visual characterization
D
E
S
I
F
X
FD=4.5
SF=8.0
SX=5.5
Image Generation
$SD=\sqrt{SF^2+FD^2}=\sqrt{8.0^2+4.5^2}=\sqrt{84.25}$
$FX=\sqrt{SF^2+SX^2}=\sqrt{8.0^2+5.5^2}=\sqrt{94.25}$
$\sin\angle SDF=\frac{SF}{SD}$
$2\cdot \mathrm{Area}_{SFD}=SD\times FE=SF\times FD=\ldots$
$\mathrm{Perimeter}_{SFD}=\ldots$
$\mathrm{Radius}_{I}=\ldots$
$EF\approx 3.9$
$\mathrm{Area}_{SFD}=\ldots$
The reasoning is successful!
Description (Text Lite version)
CoT and Q&A Generation
DATASETS

Description: Triangle SFD is a right triangle with the right angle at SFD. FD=4.5,SF=8.0. A right-angled triangle FSX with right angle at FSX.SX=5.5. Circle I is the incircle of triangle FSX. E is where the perpendicular from F meets line SD.

Step 1: Since F is 90 degrees, the length of the SD can be calculated using the Pythagorean theorem: $SD=\sqrt{SF^2+FD^2}=\sqrt{8^2+4.5^2}=\sqrt{84.25}$
Step 2: The area of triangle SFD is equal to $\frac{1}{2}\times SF\times FD=\frac{1}{2}\times 8.0\times 4.5=18.0$
Step 3:Since the area of triangle SFD equals to 18.0 SD=$\sqrt{84.25}$,we can calculate that the length of line SD equals to 2 * area of triangle SFD divided the length of SD, which finally get **3.9**.

Question:What is the length of line EF? **Choices**: A.3.9 B.5.2 C.1.7 D.2.3.
Correct letter: A

Figure 3. The overview of our neuro-symbolic data generation framework. The framework comprises three steps: In the first step, we get a symbolic language sequence in a limited symbolic action space. In the second step, the conversion engine parses the Geo-DSL sequence and translates it back to natural language and visual image without losing soundness. In the third step, we employ LLMs to take a reverse search and forward validation process to get final Q&A pairs with CoT.

hensive evaluation of the geometric reasoning capabilities of current mainstream open-source and closed-source models, with details presented in Appendix E. Notably, our training dataset consistently and efficiently enhances the geometric reasoning performance of MLLMs across multiple benchmarks. With only 4k samples and two epochs of RL training, base models achieve performance improvements of up to **+15.8%** on MathVision, **+8.4%** on MathVerse, and **+7.3%** on GeoQA. Moreover, InternVL2.5-4B can be improved to outperform the 8B model in the same series on geometric reasoning tasks.

In summary, our contributions are as follows:

- We propose **NeSyGeo**, a novel framework for geometric reasoning data generation, featuring a Geo-DSL for symbolic synthesis, a conversion engine for image and text generation, and an LLM-driven generator for Q&A pairs with CoT. NeSyGeo ensures validity through rigorous symbolic definitions and diversity via varied actions and neural searching.
- Using our framework, we synthesize the **NeSyGeo-Caption** and **NeSyGeo-CoT** training datasets with 100k high-quality samples, alongside a comprehensive geometric task evaluation set **NeSyGeo-Test**. These datasets are characterized by their diversity, rigor, and balanced distribution of information across image and text modalities.
- We demonstrate significant performance improvements on several MLLMs across multiple benchmarks using both RL and SFT training methods with our training sets, validating the effectiveness of our framework and the high quality of our datasets.

2. Related Works

2.1. Geometric Problem-Solving

Early approaches to geometric reasoning predominantly relied on symbolic solvers that used formal languages to tackle the tasks. For instance, Inter-GPS (Lu et al., 2021) and PGDP (Zhang et al., 2022) employed symbolic methods by manually crafting reasoning rules and symbolic representations for geometric entities. These systems typically transform visual input into symbolic forms through instance segmentation and apply theorem search to derive solutions. However, these methods lacked scalability due to their dependence on manually designed rules. Their inability to generalize beyond specific problem types further limited their universality and effectiveness across diverse geometric challenges.

The advent of MLLMs has shifted the paradigm toward data-driven geometric reasoning, leveraging their robust reasoning capabilities. Recent advancements include Geo-DRL (Peng et al., 2023)and GeoGen (Krueger et al., 2021). Despite these developments, geometric reasoning poses significant challenges for MLLMs, requiring seamless integration of image perception, geometric knowledge, and multi-step reasoning. GeoSense (Xu et al., 2025) identifies the identification and application of geometric principles as a persistent bottleneck. Similarly, GeoEval (Zhang et al., 2024a) reveals that current MLLMs exhibit significantly low accuracy when facing more challenging geometric problems. MathVerse (Zhang et al., 2024c) further highlights MLLMs' over-reliance on textual information, underscoring the critical need for balanced multimodal datasets to enhance cross-modal reasoning capabilities.

2.2. Multimodal Geometry Datasets

Large-scale, high-quality datasets are essential for enhancing the performance of MLLMs in solving geometric problems. Early datasets such as GeoS (Seo et al., 2015) (186 problems), Geometry3k (Lu et al., 2021) (3000 problems) and GeoQA (Chen et al., 2021) (4998 problems) utilized human manual annotation. Their datasets are thus limited to a small scale. With the development of MLLMs, datasets of greater magnitude have become essential. To address this, numerous efforts have shifted toward automatic data generation.

G-LLaVA (Gao et al., 2025) rephrased questions from GeoQA and Geometry3k to create 115,000 Q&A pairs, but failed to enhance image variety. Template-based methods (Zhang et al., 2024d; Deng et al., 2024) typically rely on 10–20 predefined geometric figures, limiting the diversity of the generated images. AlphaGeometry (Trinh et al., 2024), a notable work that combines symbolic solvers for geometric proofs, employs a symbolic language definition. Yet, due to the absence of numerical attributes such as angle measures and segment lengths in its geometric space, attempts to automatically generate datasets using the AlphaGeometry framework (Huang et al., 2025; Zhang et al., 2024b) are confined to caption datasets, failing to produce the numerical Q&A pairs critical for current MLLMs training. In contrast to prior approaches, our method pioneers a neuro-symbolic framework, being the first to integrate the precision of symbolic definition with the diversity of neural search for generating multimodal reasoning data.

3. Methods

To address the urgent need for large-scale and high-quality multimodal datasets in MLLMs for geometric reasoning, we propose **NeSyGeo**, a novel three-stage data generation pipeline. The pipeline is built upon **Geo-DSL**, a symbolic DSL designed to represent most elements in plane geometry space concisely and wholly. Its entity-relation-constraint structure allows any element to be defined via a single statement, while its expressive power ensures comprehensive coverage of all geometric elements and values.

NeSyGeo's generation process unfolds in three distinct stages: **First**, a symbolic generator performs action augmentations within the finite symbolic space to produce a Geo-DSL sequence. This approach effectively constrains the synthesis and augmentation process, ensuring validity and controllability by generating outside the infinite domains of natural language and image spaces (Section 3.2). **Second**, a conversion engine maps the generated Geo-DSL sequences back into natural language descriptions and visual image representations. This process synthesizes high-quality images and valid text while avoiding intermodal information overlap. **Third**, to get Q&A Pairs with reasoning paths, we utilize expert LLMs to conduct backwards search to identify the geometric unknowns to be solved and generate the CoT in a forward manner (Section 3.4). The search process primarily ensures the diversity of the Q&A pairs, while the forward verification confirms the correctness of the CoT and the final answer. The overall framework of NeSyGeo is illustrated in Figure 3.

3.1. Symbolic Definition

Existing symbolic languages for plane geometry have certain limitations. The definitions in AlphaGeometry (Trinh et al., 2024) are tailored for proof-based problems, thus lacking definitions related to specific numerical values. InterGPS (Lu et al., 2021) defines a predicate as a geometric shape entity, geometric relation, or arithmetic function, constructing 91 predicates. However, this approach overly fragments shape, attribute, and relation definitions into independent statements, often requiring multiple statements to specify a single element in the figure. This significantly increases the complexity of a conversion engine's identification of elements within the symbolic space and further conversion of them.

We propose Geo-DSL, a concise and comprehensive symbolic language for plane geometry to address these limitations. It employs an entity-relation-constraint framework, using well-defined rules to uniquely define 13 types of points, 7 types of lines, 3 types of angles, and 14 types of shapes, covering all plane geometry elements and incorporating numerical attributes like lengths and angles for precise specifications. Table 1 provides partial examples of symbolic definitions. Geo-DSL offers two key advantages. First, its comprehensive coverage includes all geometric elements and their numerical properties, enabling accurate and complete descriptions. Second, its simplicity allows a

Table 1. Examples of our Geo-DSL and Corresponding Natural Language. Geo-DSL is defined by an entity-relation-constraint framework, encompassing 13 point types, 7 line types, 3 angle types, and 14 shape types in plane geometry. See Appendix H for complete Geo-DSL definitions.

Type	Geo-DSL Language	Natural Language
Shape	$Triangle(A, B, C) = (x, y, \alpha)$ $Circle(O) = (x)$	Triangle ABC has $AB = x$, $BC = y$, $\angle B = \alpha$ Circle O has radius x
Point	$Foot(D, A, Line(B, C))$ $Intersection(E, Line(A, B), Line(C, D))$	D is the foot of the perpendicular from A to BC E is the intersection of line AB and line CD
Line	$Para(Line(A, B), Line(C, D), x)$	Line AB is parallel to CD, $AB = x$
Angle	$Angle(P, Q, R) = \alpha$	$\angle PQR = \alpha$

single statement to specify an element uniquely, promoting the incremental integration of the symbolic action space and the sequential parsing of statements by the conversion engine. This combination of completeness and simplicity makes Geo-DSL an efficient and powerful geometric representation and processing solution.

3.2. Symbolic Sequence Generation

To generate a Geo-DSL sequence, we introduce a step-action augmenter that iteratively synthesizes a sequence of statements, as detailed in Algorithm 1. Based on dataset preferences, we first configure the step count N, weight matrices I and A for selecting elements and actions with respective probabilities, and ranges $[l_{\text{min}}, l_{\text{max}}]$ for lengths and $[\theta_{\text{min}}, \theta_{\text{max}}]$ for angles. Then, the augmenter iteratively generates symbolic statements over N steps. For each step, we randomly sample parameters x, y, z and α, select an element $v_j \in f_v$ using weights from I, and choose an action a_k based on weights from A (see Appendix I for action details). The new statement s_{new} is then incorporated into the sequence f_s. Leveraging Geo-DSL's symbolic definitions and well-defined actions, this approach ensures the validity and accuracy of each statement. Meanwhile, randomized elements, diverse action selections, and customizable hyperparameter preferences promote diversity.

3.3. Informalization

Following the generation of sequences within the formal symbolic space, it is necessary to map them back to the text and image spaces for further processing and visualization. Our approach generates high-quality images and rigorous natural language, allocating information between them to compel MLLMs to leverage visual data effectively.

Visual image. For the visual space, our visualization engine parses each Geo-DSL statement as a geometric element to generate high-quality images using Matplotlib. The rigorous symbolic language space ensures the determinacy and uniqueness of the conversion process, thereby enabling precise image generation. Additionally, the engine produces images with detailed annotations absent from the text, requiring models to leverage images during problem-solving, enhancing their visual perception and ability to extract image-based information.

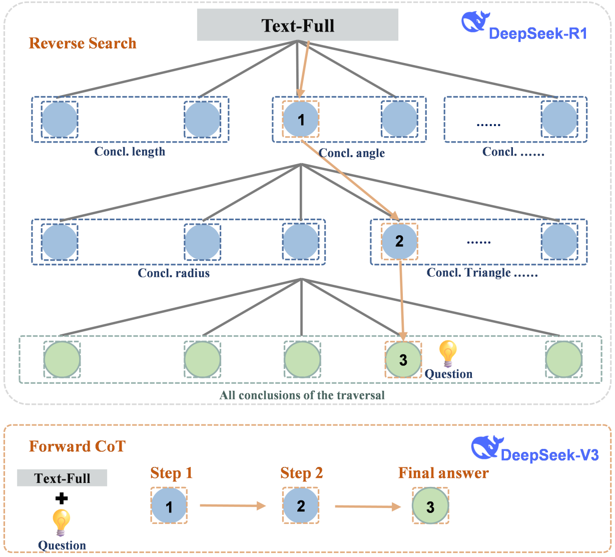


Figure 4. Reverse search and forward validation with expert LLMs

Natural language. We adopt a template-based transformation approach, parsing and mapping each symbolic statement to multiple predefined natural language templates. This produces a text-full version containing all details and captions of the image and a text-lite version retaining only essential conditions, not included in the image annotations, as the final form in our datasets. The text-lite version avoids intermodal information redundancy, while diverse templates ensure the validity and diversity of the synthesized condition text.

3.4. CoT Generation

To generate Q&A pairs with CoT reasoning, we utilize the strong reasoning abilities of expert LLMs to ensure diverse and reliable output. Using the text-full version generated as described in Section 3.3 as input, we develop a two-step process comprising reverse search and forward validation

Algorithm 1 The overall framework of the symbolic sequence generation process

```
Input: Step count $N$, Action weight matrix $A$, Element selection weight matrix $I$, Line length range $[l_{min}, l_{max}]$,
Angle range $[\theta_{min}, \theta_{max}]$.    ▷ Set customizable hyperparameter
Output: Generated Geo-DSL statement sequence $f_s$.
Initialize $f_s$=Initialize( ) .    ▷ Initialize the sequence $f_s$ with the first statement
Initialize symbolic state space elements $f_v$ = Initialize($f_s$).    ▷ Initialize state based on $f_s$
for $i = 1$ to $N$ do
    Randomly sample $x, y, z$ from $[l_{min}, l_{max}]$.
    Randomly sample $\alpha$ from $[\theta_{min}, \theta_{max}]$.
    Element $v_j$=Selected_elements($f_v$,$I$).    ▷ Select element $v_j$ from $f_v$ randomly.
    Action $a_k$=Selected_action($v_j$,$A$).    ▷ Select action $a_k$ based on the type of $v_j$ randomly
    $s_{new}$=Generate_DSL($a_k$, x, y, z, $\alpha$).    ▷ Generate new DSL statement
    $f_s$=Update($f_s$,$s_{new}$)    ▷ Add the new statement to the sequence
    $f_v$=Update($f_v$,$s_{new}$).    ▷ Update state space elements
end for
return $f_s$.
```

as shown in Figure 4. Prompts are detailed in Appendix G.

Reverse search. We use DeepSeek R1 (DeepSeek-AI et al., 2025) for reverse search, starting from the given conditions, iteratively exploring and deriving conclusions step by step, and ultimately producing the final conclusions at the end of the reasoning chain along with their corresponding answers. This reverse approach starts from known information and recursively builds a reasoning chain, reducing Q&A generation complexity and hallucinations. R1's strong reasoning and exploration capabilities yield diverse conclusions, enriching the variety of Q&A pairs.

Forward validation. To ensure the correctness of Q&A pairs and generate step-by-step CoT reasoning, we re-input the questions and text-full conditions into DeepSeek V3, requesting both the reasoning process with final answer, and cross-validate these answers against those from R1 to include only consistent pairs in our final dataset. This process guarantees answer correctness and yields diverse, valid CoT without an extensive search space.

4. Experiments

We present a series of experiments to investigate the following four research questions:

Efficacy – To what extent does training on our synthesized dataset in both RL and SFT improve the geometric reasoning performances of several MLLMs?

Efficiency – Is the data generated by NeSyGeo more effective in yielding models with better performance compared to using data from existing automatic synthesis frameworks?

Diversity – Does the NeSyGeo framework effectively ensure diversity across both the text and image spaces of the generated synthetic geometric dataset?

Visual Effectiveness – Can our datasets compel models to effectively utilize visual information for enhanced understanding by appropriately distributing information between modalities?

4.1. Experimental Setup

Dataset. To synthesize a diverse dataset, we set hyperparameters for generating images by configuring the step count, length range, and angle range. The step count N ranges from one to four. The line length is defined within the basic range $[l_{\min}, l_{\max}] = [1, 5]$, scalable by any multiple of 2 (e.g., $[4, 20]$). The angle is constrained to multiples of 15° within $[15°, 165°]$, with increased weights assigned to special angles. We generate various types of weight matrices A and I by adjusting their corresponding values. This process yields a NeSyGeo-CoT dataset with 30k Q&A pairs and a NeSyGeo-Caption dataset with 70k Q&A pairs. Additional dataset statistics are provided in Appendix B.

Evaluation. Our evaluation is conducted on several benchmarks: the Test set of GeoQA(Chen et al., 2021), the Test_MINI set of MathVision(Wang et al., 2024b), and the MathVerse(Zhang et al., 2024c). For the MathVerse benchmark, we select the Vision Only, Vision Dominant, and Vision Intensive sets to better assess the visual perception and logical reasoning capabilities of MLLMs. We extract in-domain metrics from other datasets, including angle, area, length, and Plane Geometry, to effectively evaluate the models' capabilities in geometric reasoning problems, in addition to the GeoQA dataset, which focuses entirely on plane geometry. For GeoQA, we employed hard-coded extraction for comparison, while other evaluations are assessed using the automated VLMEvalKit framework(Duan et al., 2024). Appendix C provides additional experimental details and evaluation results.

4.2. Empirical Results

Efficacy: Training multiple MLLMs with NeSyGeo via both SFT and RL significantly enhances geometric problem-solving performances.

We first sample 4k samples from our NeSyGeo-CoT dataset and apply the Group Relative Policy Optimization (GRPO) algorithm to train two epochs with Deepseek R1's format and answer rewards. The training code framework is based on VLM-R1 (Shen et al., 2025). As shown in Tables 2, models achieve the best performance among the baselines after training. InternVL2.5-4B significantly improved in the angle knowledge domain with gains of 8.4 (MathVerse), 7.3 (GeoQA), and 5.3 (MathVision). Qwen2.5-VL-3B achieved a +15.8 performance boost in the area domain of MathVision. Notably, across all evaluated metrics, the InternVL2.5-4B model trained on the NeSyGeo dataset achieves performance on par with or superior to its 8B counterpart.

We also conducted SFT experiments, initially training on our NeSyGeo-Caption dataset to enhance the models' perception of geometric images, followed by training on the NeSyGeo-CoT dataset to improve reasoning capabilities. The experiments were conducted on LLaMA-Factory (Zheng et al., 2024) framework. Evaluation results on MathVerse (Vision Intensive) and GeoQA are presented in Table 3. The trained model demonstrates performance improvements over the base model on most metrics.

Efficiency: Under the same data budget, the generated data from our framework is better than that from popular automatic generation frameworks.

We randomly sampled 4k samples from MAVIS (Zhang et al., 2024d) and R-CoT (Deng et al., 2024), which are automatic frameworks in geometry problem generation. To ensure a fair comparison, we maintained consistent settings.

Table 3. **SFT performance comparison**: The trained model demonstrates performance improvements over the base model on most metrics.

Model	GeoQA	MathVerse			
		Angle	Area	Length	Plane
Qwen2.5-VL-7B	69.4	43.0	27.5	46.2	44.1
Qwen2.5-VL-7B+Ours	**71.8** (+2.4)	**46.1** (+3.1)	**23.1** (-4.4)	**49.5** (+3.3)	**46.7** (+2.6)
LLaVA-7B	22.6	28.5	6.6	16.5	20.4
LLaVA-7B+Ours	**26.1** (+3.5)	**30.6** (+2.1)	**7.7** (+1.1)	**19.2** (+2.7)	**22.9** (+2.5)

As illustrated in Figure 5 and Table 2, while all datasets help the model improve over the baseline, our dataset outperformed others in most metrics, validating its superior performance.

Diversity: Datasets synthesized by our NeSyGeo framework exhibit high diversity in text and visual features.

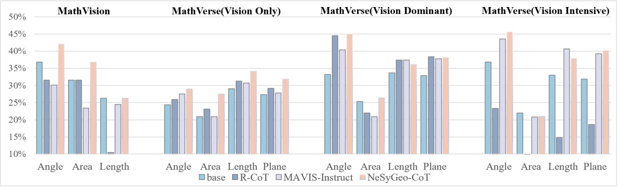


Figure 5. Efficiency comparison of our NeSyGeo-CoT dataset versus other mainstream automated synthesis datasets. The models are trained using RL methods with InternVL2.5-4B.

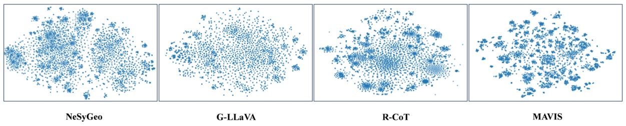


Figure 6. T-SNE of the text features of different automatic frameworks. The G-LLaVA method augments the text space on the manually annotated GeoQA dataset. Thus, its text diversity can approximate that of real data more closely. Similar to G-LLaVA, our method exhibits a uniform distribution in the space, demonstrating superior diversity.

A critical challenge for automatic data synthesis methods is whether the dataset is sufficiently diverse to avoid quality degradation due to potential overfitting risks from inherent domain constraints. We employ t-SNE (van der Maaten & Hinton, 2008) dimensionality reduction for mapping in text space to evaluate the diversity across different methods. This analysis allows us to assess the diversity of the textual descriptions themselves. Given that in geometric problems, the text conditions depict specific visual elements, and the diversity observed in the text space also serves as a valuable indicator of the diversity in the corresponding visual diagrams. To ensure a fair comparison, we remove all prompts related to guiding large models, retaining only condition and question texts, and randomly sample 5k texts from each dataset. The results are illustrated in Figure 6. Our method and G-LLaVA (Gao et al., 2025) exhibit uniformly distributed features in the space, indicating low data overlap and high diversity. In contrast, R-CoT and MAVIS display varying degrees of clustered distribution, indicating more feature-similar samples. To directly assess the diversity of the visual features, we also performed t-SNE on image features extracted by ResNet, with detailed experimental results presented in Appendix C.

Visual Perception: Models trained on NeSyGeo data shows modest gains over information-redundant datasets when textual shortcuts exist, yet achieves substantial improvements when image understanding is necessary. This indicates that ours enhances not only logical reasoning but also image perception and utilization of the model.

Table 2. **RL performance comparison**: Models trained with only 4k samples of **NeSyGeo-CoT** show performance gains over the base models, with the InternVL2.5-4B model exceeding the 8B variant in geometry problem-solving.

	GeoQA	MathVision			MathVerse			
Model		Angle	Area	Length	Angle	Area	Length	Plane
Qwen2.5-VL-3B	53.3	26.3	26.3	21.1	31.3	20.9	37.0	32.5
Qwen2.5-VL-3B+Ours	**55.7** (+2.4)	**26.3** (+0.0)	**42.1** (+15.8)	**26.3** (+5.2)	**32.6** (+1.3)	**23.5** (+2.6)	**37.2** (+0.2)	**35.5** (+3.0)
InternVL2.5-4B	61.9	36.8	31.6	26.3	31.5	22.7	31.9	30.7
InternVL2.5-4B+MAVIS	63.5 (+1.6)	31.6 (-5.2)	26.3 (-5.3)	31.6 (+5.3)	37.1 (+5.6)	20.9 (-1.8)	35.3 (+3.4)	33.7 (+3.0)
InternVL2.5-4B+R-CoT	63.3 (+1.4)	31.6 (-5.2)	31.6 (+0.0)	21.1 (-5.2)	31.2 (-0.3)	18.3 (-4.4)	34.3 (+2.4)	28.7 (-2.0)
InternVL2.5-4B+Ours	**69.2** (+7.3)	**42.1** (+5.3)	**36.8** (+5.2)	**26.3** (+0.0)	**39.9** (+8.4)	**24.9** (+2.2)	**36.1** (+4.2)	**36.7** (+6.0)
InternVL2.5-8B	66.2	36.8	36.8	21.1	36.9	23.1	34.8	36.6

Table 4. Comparison between NeSyGeo-CoT and text-redundant datasets with equivalent data budgets. Models are trained via RL on the InternVL2.5-4B. The highest value for each metric is underlined. The results show that our dataset improves models' visual perception and logical reasoning capabilities.

	Text Dominant				Vision Only			
Dataset	Angle	Area	Length	Plane Geometry	Angle	Area	Length	Plane Geometry
Base	47.1	27.5	43.4	44.1	24.4	20.9	29.1	27.3
NeSyGeo	49.2	28.6	44.0	45.5	29.0	27.5	34.1	31.8
NeSyGeo+RED	52.8	27.5	44.5	45.1	27.5	25.3	33.0	30.2
R-CoT	51.8	24.2	45.0	46.5	25.9	23.1	31.3	29.2

A key question is whether reducing redundancy and forcing models to extract visual information improves their geometric reasoning capabilities. We evaluate models on the MathVerse(Text Dominant), which provides redundant text descriptions and implicit properties enabling reasoning without images, and the MathVerse(Vision Only) version, where all information is embedded entirely within the images. For this comparison, we selected two text-redundant datasets: NeSyGeo+RED, the original NeSyGeo-CoT dataset supplemented with textual equivalents of its image annotations, and the R-CoT dataset. Results presented in Table 4 show that our model outperforms the baseline on Text-Dominant but lags behind other datasets in some metrics. On Vision-Only, our model surpasses them across all metrics, demonstrating enhanced geometric reasoning and visual perception.

5. Conclusion

This paper introduces NeSyGeo, a neurosymbolic framework for automatically synthesizing multimodal geometric datasets. Our approach transforms the generation process into a controllable symbolic space using Geo-DSL, maps the symbolic representation back to image and natural language spaces via a conversion engine, and then utilizes LLMs for backwards search and forward solving to produce Q&A pairs. Using this framework, we construct the NeSyGeo-CoT and NeSyGeo-Caption datasets, totalling 100k samples. We also propose NeSyGeo-Test, a comprehensive benchmark for evaluating MLLMs' geometric reasoning capabilities. Our datasets significantly and consistently improve the reasoning abilities of multiple MLLMs through both SFT and RL.

Future Work: We intend to extend NeSyGeo to other multimodal domains, such as analytical geometry and visual question answering. This extensibility will be achieved by defining new domain-specific languages, corresponding synthesis rules within the symbolic space, and tailored conversion engines. Furthermore, we plan to develop an automated symbolic solver capable of conducting search and validation directly within the symbolic space. This would remove reliance on LLMs, potentially reducing generation costs and ensuring complete correctness of the datasets.

6. Acknowledgements

This research is supported by the National Science Foundation of China (62306133), and the Key Program of Jiangsu Science Foundation (BK20243012). We would like to thank reviewers for their constructive suggestions.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., et al. Flamingo: a visual language model for few-shot learning. In *Advances in Neural Information Processing Systems*, pp. 23716–23736, 2022.

Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, S., Tang, J., et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025.

Chen, J., Tang, J., Qin, J., Liang, X., Liu, L., Xing, E. P., and Lin, L. Geoqa: A geometric question answering benchmark towards multimodal numerical reasoning. In *Findings of the Association for Computational Linguistics*, pp. 513–523, 2021.

Chen, Z., Wu, J., Wang, W., Su, W., Chen, G., Xing, S., Zhong, M., Zhang, Q., Zhu, X., Lu, L., et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24185–24198, 2024.

DeepSeek-AI, Guo, D., et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Deng, L., Liu, Y., Li, B., Luo, D., Wu, L., Zhang, C., Lyu, P., Zhang, Z., Zhang, G., Ding, E., et al. R-cot: Reverse chain-of-thought problem generation for geometric reasoning in large multimodal models. *arXiv preprint arXiv:2410.17885*, 2024.

Duan, H., Yang, J., Qiao, Y., Fang, X., Chen, L., Liu, Y., Dong, X., Zang, Y., Zhang, P., Wang, J., et al. Vlmevalkit: An open-source toolkit for evaluating large multi-modality models. In *Proceedings of the ACM International Conference on Multimedia*, pp. 11198–11201, 2024.

Gao, J., Pi, R., Zhang, J., Ye, J., Zhong, W., Wang, Y., Hong, L., Han, J., Xu, H., Li, Z., and Kong, L. G-llava: Solving geometric problem with multi-modal large language model. In *The 13th International Conference on Learning Representations*, 2025.

Huang, Z., Wu, T., Lin, W., Zhang, S., Chen, J., and Wu, F. Autogeo: Automating geometric image dataset creation for enhanced geometry understanding. *IEEE Transactions on Multimedia*, 2025.

Jiang, Y., Zhang, J., Sun, K., Sourati, Z., Ahrabian, K., Ma, K., Ilievski, F., and Pujara, J. Marvel: Multidimensional abstraction and reasoning through visual evaluation and learning. In *Advances in Neural Information Processing Systems*, pp. 46567–46592, 2024.

Kazemi, M., Alvari, H., Anand, A., Wu, J., Chen, X., and Soricut, R. Geomverse: A systematic evaluation of large models for geometric reasoning. In *AI for Math Workshop at the International Conference on Machine Learning*, 2024.

Krueger, R., Han, J. M., and Selsam, D. Automatically building diagrams for olympiad geometry problems. In *International Conference on Automated Deduction*, pp. 577–588, 2021.

Li, J., Selvaraju, R., Gotmare, A., Joty, S., Xiong, C., and Hoi, S. C. H. Align before fuse: Vision and language representation learning with momentum distillation. In *Advances in Neural Information Processing Systems*, pp. 9694–9705, 2021.

Liang, Y., Cai, Z., Xu, J., Huang, G., Wang, Y., Liang, X., Liu, J., Li, Z., Wang, J., and Huang, S.-L. Unleashing region understanding in intermediate layers for MLLM-based referring expression generation. In *Advances in Neural Information Processing Systems*, pp. 120578–120601, 2024.

Lin, W., Wei, X., An, R., Gao, P., Zou, B., Luo, Y., Huang, S., Zhang, S., and Li, H. Draw-and-understand: Leveraging visual prompts to enable MLLMs to comprehend what you want. In *The 13th International Conference on Learning Representations*, 2025.

Liu, H., Li, C., Wu, Q., and Lee, Y. J. Visual instruction tuning. In *Advances in Neural Information Processing Systems*, pp. 34892–34916, 2023.

Lu, P., Gong, R., Jiang, S., Qiu, L., Huang, S., Liang, X., and Zhu, S.-C. Inter-gps: Interpretable geometry problem solving with formal language and symbolic reasoning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, pp. 6774–6786, 2021.

Peng, S., Fu, D., Liang, Y., Gao, L., and Tang, Z. Geodrl: A self-learning framework for geometry problem solving using reinforcement learning in deductive reasoning. In *Findings of the Association for Computational Linguistics*, pp. 13468–13480, 2023.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Trogden, W., and Sutskever, I. Learning transferable visual models from natural language supervision. In *Proceddings of the 38th International Conference on Machine Learning*, pp. 8748–8763, 2021.

Seo, M., Hajishirzi, H., Farhadi, A., Etzioni, O., and Malcolm, C. Solving geometry problems: Combining text and diagram interpretation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1466–1476, 2015.

Shen, H., Liu, P., Li, J., Fang, C., Ma, Y., Liao, J., Shen, Q., Zhang, Z., Zhao, K., Zhang, Q., Xu, R., and Zhao, T. Vlm-r1: A stable and generalizable r1-style large vision-language model. *arXiv preprint arXiv:2504.07615*, 2025.

Trinh, T. H., Wu, Y., Le, Q. V., He, H., and Luong, T. Solving olympiad geometry without human demonstrations. *Nature*, pp. 802–809, 2024.

van der Maaten, L. and Hinton, G. Visualizing data using t-sne. In *Journal of Machine Learning Research*, pp. 2579–2605, 2008.

Wang, A. J., Li, L., Lin, Y., Li, M., Wang, L., and Shou, M. Z. Leveraging visual tokens for extended text contexts in multi-modal learning. In *Advances in Neural Information Processing Systems*, pp. 14325–14348, 2024a.

Wang, K., Pan, J., Shi, W., Lu, Z., Ren, H., Zhou, A., Zhan, M., and Li, H. Measuring multimodal mathematical reasoning with math-vision dataset. In *Advances in Neural Information Processing Systems*, pp. 95095–95169, 2024b.

Wang, P., Bai, S., Tan, S., Wang, S., Fan, Z., Bai, J., Chen, K., Liu, X., Wang, J., Ge, W., et al. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*, 2024c.

Wu, M., Cai, X., Ji, J., Li, J., Huang, O., Luo, G., Fei, H., Jiang, G., Sun, X., and Ji, R. ControlMLLM: Training-free visual prompt learning for multimodal large language models. In *Advances in Neural Information Processing Systems*, pp. 45206–45234, 2024.

Xu, L., Zhao, Y., Wang, J., Wang, Y., Pi, B., Wang, C., Zhang, M., Gu, J., Li, X., Zhu, X., et al. Geosense: Evaluating identification and application of geometric principles in multimodal reasoning. *arXiv preprint arXiv:2504.12597*, 2025.

Yan, Y., Su, J., He, J., Fu, F., Zheng, X., Lyu, Y., Wang, K., Wang, S., Wen, Q., and Hu, X. A survey of mathematical reasoning in the era of multimodal large language model: Benchmark, method & challenges. *arXiv preprint arXiv:2412.11936*, 2024.

Yan, Y., Wang, S., Huo, J., Ye, J., Chu, Z., Hu, X., Yu, P. S., Gomes, C., Selman, B., and Wen, Q. Position: Multimodal large language models can significantly advance scientific reasoning. *arXiv preprint arXiv:2502.02871*, 2025.

Zhang, J., Li, Z.-Z., Zhang, M.-L., Yin, F., Liu, C.-L., and Moshfeghi, Y. Geoeval: Benchmark for evaluating LLMs and multi-modal models on geometry problem-solving. In *Findings of the Association for Computational Linguistics*, pp. 1258–1276, 2024a.

Zhang, J., Liu, O., Yu, T., Hu, J., and Neiswanger, W. Euclid: Supercharging multimodal llms with synthetic high-fidelity visual descriptions. *arXiv preprint arXiv:2412.08737*, 2024b.

Zhang, J., Khayatkhoei, M., Chhikara, P., and Ilievski, F. MLLMs know where to look: Training-free perception of small visual details with multimodal LLMs. *arXiv preprint arXiv:2502.17422*, 2025a.

Zhang, M.-L., Yin, F., Hao, Y.-H., and Liu, C.-L. Plane geometry diagram parsing. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence*, pp. 1636–1643, 2022.

Zhang, R., Jiang, D., Zhang, Y., Lin, H., Guo, Z., Qiu, P., Zhou, A., Lu, P., Chang, K.-W., Gao, P., and Li, H. Mathverse: Does your multi-modal llm truly see the diagrams in visual math problems? In *European Conference on Computer Vision*, pp. 169–186, 2024c.

Zhang, R., Wei, X., Jiang, D., Zhang, Y., Guo, Z., Tong, C., Liu, J., Zhou, A., Wei, B., Zhang, S., et al. Mavis: Mathematical visual instruction tuning. *arXiv preprint arXiv:2407.08739*, 2024d.

Zhang, X., Zhu, N., Qin, C., Li, Y., Zeng, Z., and Leng, T. Fgeo-hypergnet: Geometric problem solving integrating formal symbolic system and hypergraph neural network. In *Proceedings of the 34rd International Joint Conference on Artificial Intelligence*, 2025b.

Zhao, J., Zhang, T., Sun, J., Tian, M., and Huang, H. Pi-gps: Enhancing geometry problem solving by unleashing the power of diagrammatic information. *arXiv preprint arXiv:2503.05543*, 2025.

Zheng, Y., Zhang, R., Zhang, J., Ye, Y., Luo, Z., Feng, Z., and Ma, Y. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pp. 400–410, 2024.

Table 5. Statistics of NeSyGeo-Caption

Statistic	Number
Total Counts	
Total number of images	70k
Total number of captions	70k
DSL Statement Percentage	
One Statement	5.4%
Two Statements	25.8%
Three Statements	34.7%
Four Statements	23.4%
Five Statements	10.7%
Length of Captions	
Maximum length (words)	220
Minimum length (words)	34
Average length (words)	73.3
Average length (characters)	385.4
Image Dimensions	
Average dimensions (pixels)	723.1 × 724.0

Table 6. Statistics of NeSyGeo-CoT

Statistic	Number
Total Counts	
Total number of images	15.3k
Total number of Q&A pairs	30.1k
Question Statistics	
Length-based type	54.6%
Area-based type	34.4%
Angle-based type	11.1%
Average length (words)	26.9
Average length (characters)	140.6
CoT Statistics	
Below four steps	35.4%
Four steps or above	64.5%
Average length (words)	91.8
Average length (characters)	365.9
Image Dimensions	
Average dimensions (pixels)	731.0 × 727.4

A. Comparison with Specific Examples of Popular Geometry Datasets

To facilitate comparison of dataset characteristics synthesized by our method and other popular approaches, we showcase a randomly selected example from NeSyGeo-CoT alongside each of the different approaches in Figure 7. Geometry-3K is a manually synthesized dataset, while the remaining approaches employ automatic generation techniques. To ensure a fair comparison, we standardize the text format by removing model-guiding prompts and appending options when present. Furthermore, we annotate each sample with image pixels and CoT word counts.

Compared to other datasets, our dataset features clear, human-aesthetically pleasing images, high-quality step-by-step reasoning chains, symbolic form meta-information enabling subsequent image augmentation and mutation, and well-distributed conditional information between images and text. Additional examples of our NeSyGeo-CoT dataset can be found in the Appendix 11.

B. Statistics of NeSyGeo-Caption and NeSyGeo-CoT

Detailed numerical statistics and element distribution for the NeSyGeo-Caption and NeSyGeo-CoT datasets are presented in Table 5 and 6. For element distribution statistics, we randomly sampled 1.8k Geo-DSL sequences corresponding to images from each dataset, counting the frequency of different geometric elements. To facilitate interpretation, these elements are converted into corresponding natural language descriptions.

C. Additional Experimental Details and Results

We utilized the VLM-R1 (Shen et al., 2025) framework for RL experiments, conducted on 6 vGPU-32 GB. We set epochs to 2, num generations to 6, batchsize to 1. To enhance the visual perception capabilities of MLLMs, parameters of the language model and vision modules are set to be trainable.

For SFT experiments, we employed the LLaMA-Factory (Zheng et al., 2024) framework on 2 A800 GPUs with LoRA. We set the learning rate of 1×10^{-5}, LoRA rank of 64, and use Adam optimization. Training on NeSyGeo-Caption used 1 epoch, while NeSyGeo-CoT used 2 epochs.

Table 7 presents the detailed performance of models trained on various automatically synthesized datasets across the MathVerse benchmark. Models trained using our dataset demonstrate superior performance on most metrics compared to others, exhibiting substantial performance gains relative to the base model.As shown in Figure 9. We also evaluated model

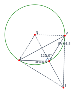
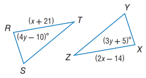
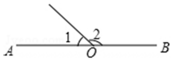
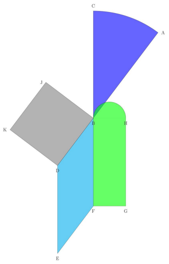
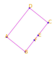
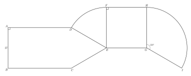
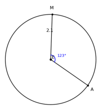
Dataset	Text	Image	Image Symbolic Form	Chain of Thought	Answer
NesyGeo-CoT	Triangle UFV. J is reflection of V over FU. Circle N is the circumcircle of triangle UFV. What is the diameter of circle N? "Choices": "A.7.5, B.8.0, C.8.5, D.9.0"	 (675 x 790)	Triangle(U,F,V)=(4.5,4.5,12), Mirror(J,V,Line(F,U)), Cir_circle(N,Triangle(U,F,V))	Step1: Triangle UFV is isosceles with UF = FV = 4.5 and ∠F = 120°.To find the diameter of the circumcircle (Circle N), we first need to find the length of UV.Step3: Using the Law of Cosines in triangle UFV… (145 words)	D
Geometry-3k	Triangle R S T cong triangle X Y Z. Find y. "choices": ["5","14","15","35"],	 R (x + 21) T (4y − 10)° S Y (3y + 5)° X Z (2x − 14) (864 x 434)	"Equals(LengthOf(Line(T, R)), x+21)", "Equals(MeasureOf(Angle(T, R, S)), 4y-10)", "Equals(MeasureOf(Angle(Z, X, Y)), 3y+5)", "Equals(LengthOf(Line(Z, X)), 2x-14)"	\	C
G-LLaVA	In the given diagram, if angle 1 has a measure of 35.0 degrees, what is the measure of angle 2? choices": A: 55° B: 135° C: 145° D: 155°	 (165 x 56)	\	Since angle 1 + angle 2 = 180°, and angle 1 = 35°, therefore angle 2 = 145°. Therefore, option C is selected. (19words)	C
Geomverse	If the arc length of the ABC sector is 12.85, the area of the BDEF parallelogram is 108, the BFGH shape is a combination of a rectangle and a semi-circle, the length of the FG side is 6, the perimeter of the BFGH shape is 48, the area of the BJKD square is 121 and the angle DBF is vertical to CBA, compute the length of the BC side of the ABC sector. Assume $\\pi=3.14$. Round computations to 2 decimal places.	 (811 x 1280)	\	The perimeter of the BFGH shape is 48 and the length of the FG side is 6, so $2 * OtherSide + 6 + \\frac{6 * 3.14}{2} = 48$. So $2 * OtherSide = 48 - 6 - \\frac{6 * 3.14}{2} = 48 - 6 - \\frac{18.84}{2} = 48 - 6 - 9.42 = 32.58$. Therefore, the length of the BF side is $\\frac{32.58}{2} = 16.29$. The area of the BJKD square is 121, so the length of the BD side is $\\sqrt{121} = 11$... (188 words)	20.4
AutoGeo	Render a clear and concise description of a image about geometric shapes.	 (1280 x 1280)	\	\	Rectangle ABCD.Point E is positioned in a way that line EB is perpendicular to point B.F lies on line segment CB.
MAVIS-Instruct	Side CD materializes as an equilateral triangle. DEF is identified as a sector. FEGH is a square. HGI is in the form of a sector. Please provide the length of arc HI in sector HGI.	 (4434 x 1596)	\	Given that AB is 27, rectangle has equal oppsite edges, so CD is also 27. Since CDE is a equilateral triangle…(88 words)	18*π
R-CoT	Give reasoning steps and answers. In the diagram, there is a circle E with radius 2.1. What is the circumference of the circle E?	 (514 x 404)	\	Step1: The circumference of a circle is calculated using the formula C = 2\u03c0r. Step2: Substituting the given value, EM = 2.1, we get C = 2\u03c0*2.1 = 4.2\u03c0. The answer is 4.2\u03c0… (34 words)	4.2

Figure 7. Comparison NeSyGeo-CoT dataset with other Popular Geometry Datasets. Geometry-3K is a manually synthesized dataset, while the remaining approaches employ automatic generation techniques. Our dataset features clear, human-aesthetically pleasing images, high-quality step-by-step reasoning chains, symbolic form meta-information enabling subsequent image augmentation and mutation, and well-distributed conditional information between images and text.

Table 7. Detailed RL experiments evaluation on MathVerse. Here, 'AGL', 'ARA', 'LTH', and 'PG' denote angle, area, length, and plane geometry, respectively.

	Vision Intensive				Vision Dominant				Vision Only			
Model	AGL	ARA	LTH	PG	AGL	ARA	LTH	PG	AGL	ARA	LTH	PG
Qwen2.5-VL-3B	31.6	22.0	34.6	33.3	31.6	17.6	40.7	31.4	30.6	23.1	35.7	32.7
InternVL2.5-4B	36.8	22.0	33.0	31.8	33.2	25.3	33.7	32.9	24.4	20.9	29.1	27.3
InternVL2.5-8B	44.0	23.1	36.3	41.8	40.4	20.9	36.8	37.3	26.4	25.3	31.3	30.6
Qwen2.5-VL-3B+RL	33.7	23.1	36.8	34.7	32.1	20.9	37.4	36.6	32.1	26.4	37.4	35.1
InternVL2.5-4B+RL	45.6	20.9	37.9	40.2	45.1	26.4	36.2	38.2	29.0	27.5	34.1	31.8

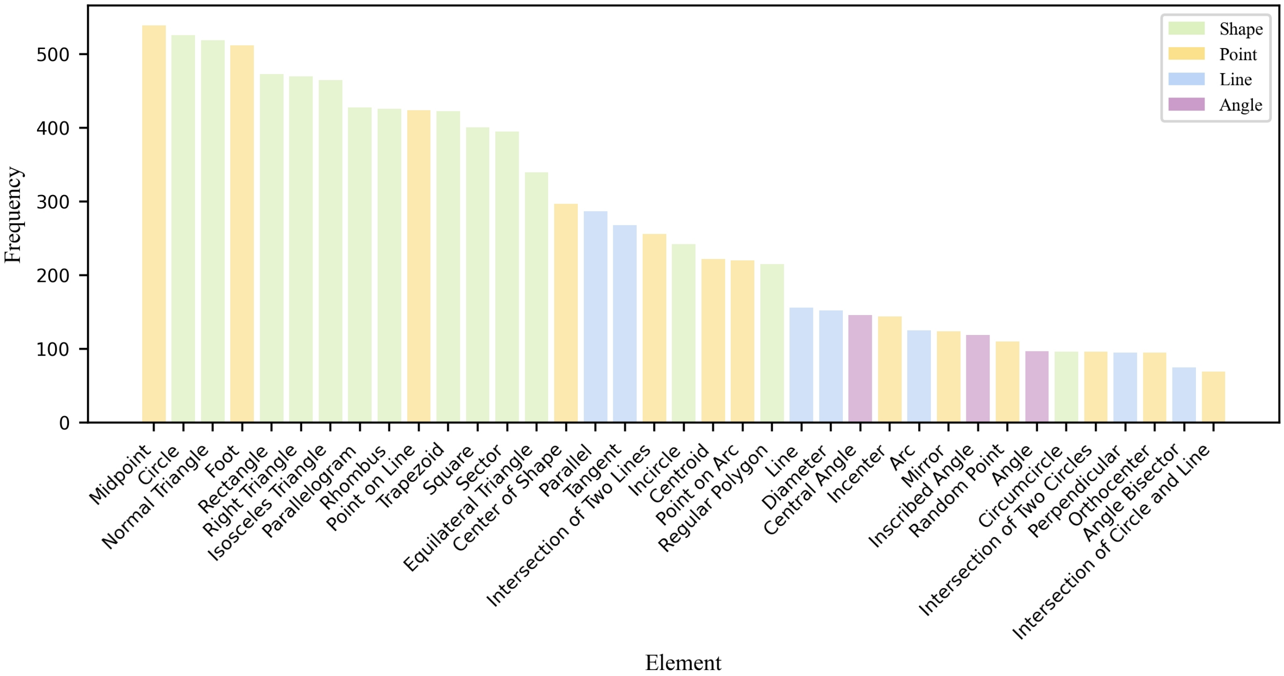


Figure 8. Frequency of different geometric elements. To facilitate interpretation, these elements are converted into corresponding natural language descriptions.

performance as RL training steps increased when using NeSyGeo-CoT. Most metrics improved with more training steps, demonstrating the robustness and effectiveness of our datasets.

To investigate visual diversity directly across datasets, we randomly sampled 1k images from each, extracted features using ResNet, and visualized them via t-SNE. As illustrated in Figure 10, G-LLaVA—having augmented only the textual components of its base dataset—displays a distinctly non-uniform distribution in the image feature space. Conversely, our method exhibits uniform feature distributions, underscoring the substantial visual diversity of images generated by our approach.

D. More Examples of NeSyGeo-CoT Dataset

We present more examples from the **NeSyGeo-CoT** dataset in Figure 11. Our bidirectional conversion engine can generate high-quality visual images from a symbolic form based on our Geo-DSL language. To further enhance image diversity, the engine introduces variability by randomly selecting values for the unit length and applying random rotations to the generated

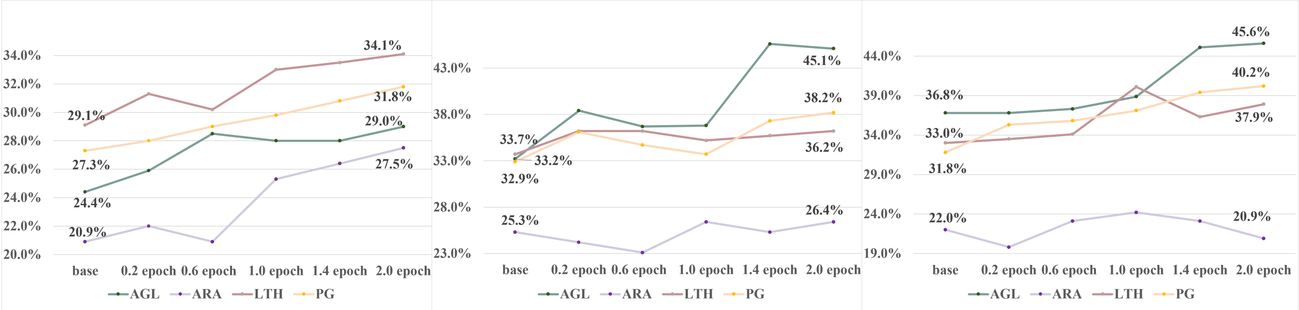


Figure 9. Model performance on Mathverse as the RL training steps increase. With InternVL2.5-4B as our base model, most metrics exhibit progressive improvement throughout training, demonstrating the robustness and effectiveness of our datasets.

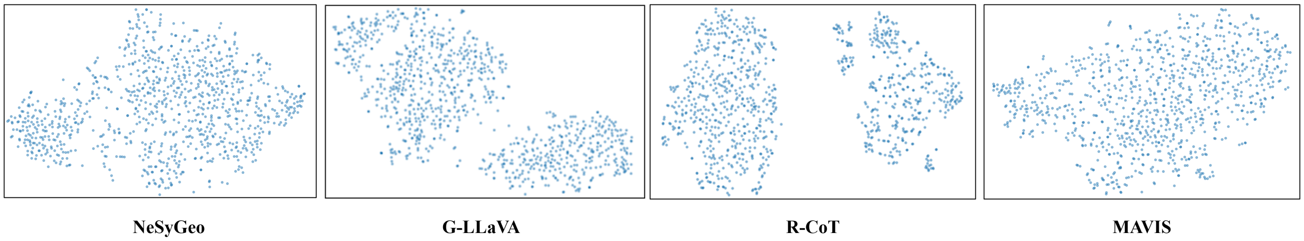


Figure 10. T-SNE of the image features of different automatic frameworks. Our datasets exhibit uniform feature distributions, underscoring the substantial visual diversity of images generated by the NeSyGeo framework.

diagrams during creation. While other visual attributes, such as element and background colours, could also be randomized, they were set to default values in our current synthesis process. Our symbolic language helps identify parts of the image, and our conversion process ensures the images are geometrically correct.

Due to LLMs' powerful search and reasoning capabilities, we obtain diverse Q&A pairs concerning properties like lengths, angles, or areas alongside high-quality CoT step-by-step. This process, involving backwards search across the geometric space defined by the symbolic form and forward validation, ensures the correctness of the numerical answers, thereby enriching textual diversity.

To support evaluation and training paradigms such as curriculum learning, we annotate each sample with a difficulty level. Given that geometric reasoning tasks primarily require models' image perception and logical reasoning capabilities, we scientifically define the difficulty level as

$$0.3 \times \text{perception difficulty} + 0.7 \times \text{reasoning difficulty}, \tag{1}$$

where perception difficulty is the number of Geo-DSL statements, and reasoning difficulty is the number of reasoning steps.

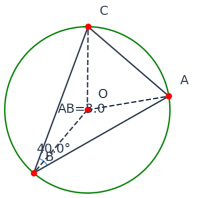
Each synthesized sample includes detailed meta-information stored as a symbolic form based on our Geo-DSL language. This symbolic form accurately describes the geometric setup and offers promising directions for future research. For instance, valid geometric configurations could be generated by augmenting or mutating existing symbolic forms within constrained parametric bounds.

E. Details of NeSyGeo-Test Benchmark.

Our NeSyGeo-Test benchmark comprises 2668 Q&A pairs. Consistent with the training set, numerical annotations are embedded in the image space, with only essential conditions and questions provided in the text. The type of numerical quantity categorizes the dataset sought: Angle (658 pairs), Shape (730 pairs), and Length (1280 pairs). Shape type includes shape area and perimeter, while length type includes edge and arc lengths. Based on the difficulty level in D, problem difficulty is divided into three levels: Easy (1537 pairs), Medium (908 pairs), and Hard (223 pairs). Evaluation results on current mainstream open-source and closed-source MLLMs are shown in Table 8.

F. Limitations

- **Limited Training Paradigm**: Our current evaluation of the NeSyGeo dataset relies on a simple training paradigm to assess its efficacy for automated data generation. This approach lacks advanced training strategies, such as CLIP alignment or curriculum learning, which restricts the development of specialized models optimized for geometric reasoning tasks.
- **Restricted Domain Scope**: The NeSyGeo framework is currently tailored to plane geometry, limiting its generalizability to other domains. However, we believe that for multimodal datasets in other domains, we can similarly achieve synthesis by defining symbolic statements, shifting the synthesis process to a controllable symbolic space, and constructing a symbolic-to-image engine, which we plan to explore in our future work.
- **Dependency on External APIs**: The construction of Q&A pairs in this study partly relies on the reasoning capabilities of LLMs. This dependency increases generation costs and introduces potential inconsistencies. We aim to develop an



Symbolic Form: Iso_triangle(A,B,C)=(3,40), Cir_circle(O,Triangle(A,B,C))

Element Types: Triangle,Circle
Difficulty Level: 3

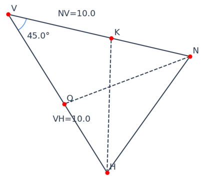
Caption : There is a isosceles triangle ABC, O is the circumcircle of ABC

Question: What is the measure of angle OCB?

Choices: A.60° B.40° C.20° D.10°

Reasoning steps: Step 1: Since angle ABC = 40° and BA = BC, angle ACB = (180° - 40°) / 2 = 70°. Step 2: The central angle subtended by the same chord is twice the inscribed angle, so angle AOC = 2 * 40° = 80°. Step 3: Since OA = OC = the radius of the circle, angle OCA = (180° - 80°) / 2 = 50°. Step 4: Therefore, angle OCB = 70° - 50° = 20°.

Answer: C.



Symbolic Form: Triangle(N,V,H)=(10.0,10.0,45), Angle_bisector(Q,Angle(H,N,V),Line(V,H)), Angle_bisector(K,Angle(V,H,N),Line(N,V))

Element Types: Triangle, Angle_bisector
Difficulty Level: 4

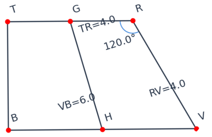
Caption : Consider triangle NVH. Angle HNV is bisected by ray Q. Line K divides angle VHN into two equal parts

Question: What is the length of side KN?

Choices: A.4.3 B.5.3 C.6.3 D.7.3

Reasoning steps: Step1: Given triangle NVH with VH = NV = 10 and angle V = 45°. Step2: The Angle Bisector Theorem states VK/KN = VH/HN. HN can be found using the Law of Cosines in triangle NVH: HN² = NV² + VH² - 2*NV*VH*cos(V) = 10² + 10² - 2*10*10*cos(45°) ≈ 200 - 141.42 ≈ 58.58, so HN ≈ 7.7. Step3: According to the Angle Bisector Theorem, VK/KN = VH/HN ≈ 10/7.7 ≈ 1.3. Step4: Let VK = 1.3x and KN = x. Then VK + KN = VN = 10, so 1.3x + x = 10 → 2.3x = 10 → x ≈ 4.3. Therefore, KN ≈ 4.3.

Answer: A.



Symbolic Form: Trapezoid(T,R,V,B)=(4,4,6,120)
IsMidpointOf(G,Line(R,T))
IsMidpointOf(H,Line(V,B))
Line(G,H)

Element Types : Trapezoid, Midpoint
Difficulty Level: 5

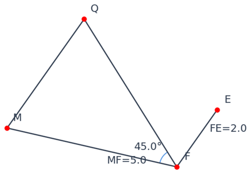
Caption :A trapezoid with vertices TRVB.G divides line segment RT into two equal parts.Point H is the midpoint of line segment VB. Line GH.

Question: What is the area of shape TGHB?

Choices: A.11.3 B.13.5 C.5.7 D.8.7

Reasoning steps: Step1: Draw trapezoid TRVB with TR∥VB, TR=4,VB=6, RV=4, ∠TRV=120°. First, Calculate height (h):Drop a perpendicular from R to VB at P. In ΔRVP: ∠RVP=60°, RV=4 → h=4×sin60°=2√3≈3.464. Step2: Find midpoints:G is midpoint of RT → GT = 2. H is midpoint of VB → VH=3. Step3: Calculate areas:Total trapezoid area = ½(4+6)×2√3=10√3≈17.32. Step4: ΔHRG area =½×2×2√3=2√3 ≈3.464.ΔVHR area =½×3×2√3=3√3≈5.196. Step5: Compute TGHB area: 17.32 - 3.464 - 5.196 ≈ 8.66 ≈ 8.7.

Answer: C.



Symbolic Form: Iso_triangle(M,F,Q)=(5,45)
Paral(Line(E,F),Line(M,Q),2)

Element Types: Triangle,
Difficulty Level: 2

Caption : Triangle MFQ is isosceles.We have line MQ parallel to line FE

Question: What is the measure of angle QFE?

Choices: A.45° B.67.5° C.73.5° D.53.5°

Reasoning steps: Step1: Triangle MFQ is isosceles with MF = FQ (given). Therefore, angle FMQ = angle MQF. angle FMQ + angle MQF = 180 - 45 = 135 degrees. Since angle FMQ = angle MQF, each angle is 135 / 2 = 67.5 degrees. Step2: Since MQ is parallel to FE and FQ is a transversal, angle MQF is equal to angle QFE by the alternate interior angles theorem, therefore angle QFE = angle MQF = 67.5 degrees.

Answer: B.

Figure 11. Examples of the **NeSyGeo-CoT** dataset. Each sample comprises a symbolic image definition based on our Geo-DSL language, a high-quality annotated image, a concise text caption, diverse Q&A pairs, and a detailed reasoning process step-by-step.

Table 8. NeSyGeo-Test Benchmark on several mainstream MLLMs. The highest accuracy for open-source and closed-source MLLMs is marked in red and blue, respectively.

		Task Type			Question Difficulty			
Model	Param	Angle	Shape	Length	Easy	Medium	Hard	Total
Open-source MLLMs								
Qwen2.5-VL-3B-Instruct	3B	44.1	31.8	27.8	30.5	29.1	31.9	30.9
Qwen2.5-VL-7B-Instruct	7B	38.3	30.7	32.2	36.8	27.5	32.3	43.3
InternVL2.5-4B	4B	53.8	53.1	53.5	62.9	49.4	32.0	53.4
InternVL2.5-8B	8B	55.9	56.2	55.4	64.5	49.4	43.3	55.8
LLaVA-NeXT-7B	7B	23.7	15.5	15.4	18.6	14.7	13.3	6.4
LLaVA-NeXT-13B	13B	15.7	15.6	16.0	15.8	16.0	15.2	15.8
LLaVA-NeXT-34B	34B	23.7	21.0	18.5	19.1	21.5	19.2	19.9
Closed-source MLLMs								
InternVL3-latest	–	81.7	65.2	68.3	77.8	62.0	56.7	68.7
GPT-4o-mini	–	58.7	62.1	55.7	63.0	54.0	41.7	58.2
Claude-3.5-Sonnet-latest	–	68.8	78.0	71.0	77.8	73.2	56.5	74.5
Qwen-VL-plus	–	38.5	29.6	31.7	36.6	27.2	29.6	32.8
Gemini-2.0-Flash	–	36.8	60.4	67.7	54.3	63.8	61.0	58.1

automated solver that conducts search and validation directly within the symbolic space, thereby removing reliance on LLMs. This could further reduce costs and ensure complete rigor.

G. Details of Prompts in Reverse Search and Forward Validation

In our automatic synthesis framework, we employ DeepSeek R1 as the expert LLM for reverse search and DeepSeek V3 for forward validation. The specific prompts utilized are detailed in Figures 12 and 13, respectively. Note that the blue text in these prompts is substituted with actual content.

H. Detailed Definition of Geo-DSL

Geo-DSL adopts an entity-relation-constraint framework to define geometric elements in plane geometry, encompassing 13 types of points, 7 types of lines, 3 types of angles, and 14 types of shapes. Representative examples of symbolic statements and their corresponding natural language descriptions are illustrated in Figures 18, 15, 17, and 16. With a single statement, Geo-DSL uniquely specifies spatial elements, ensuring the accuracy of geometric synthesis while significantly facilitating parsing and transformation by our conversion engine. This language achieves comprehensive coverage of plane geometry, including numerical attributes such as lengths and angle measures, enabling precise and complete geometric representations. By streamlining definitions into concise statements, Geo-DSL reduces the complexity of symbolic processing, enhances the efficiency of the conversion engine, and supports seamless integration with neural synthesis pipelines. These advantages make Geo-DSL a robust and versatile solution for generating high-quality multimodal geometric reasoning data.

I. Detailed Actions in Symbolic Spaces

As illustrated in Figure 14, we enumerate all statements within the action space defined by our Geo-DSL. The content within square brackets denotes annotations for each statement.

As outlined in Algorithm 1, for each step, we first generate three lengths, x, y, and z, along with an angle α, sampled from predefined ranges. Subsequently, based on the weight matrices A and I, we determine the specific statement to be selected. The chosen statement, paired with its corresponding numerical values, is then appended to the Geo-DSL sequence. For different types of actions, we provide a concrete example for each, highlighted with a gray background to indicate the available action space when the respective element is selected.

The prompt for DeepSeek R1 to reverse search

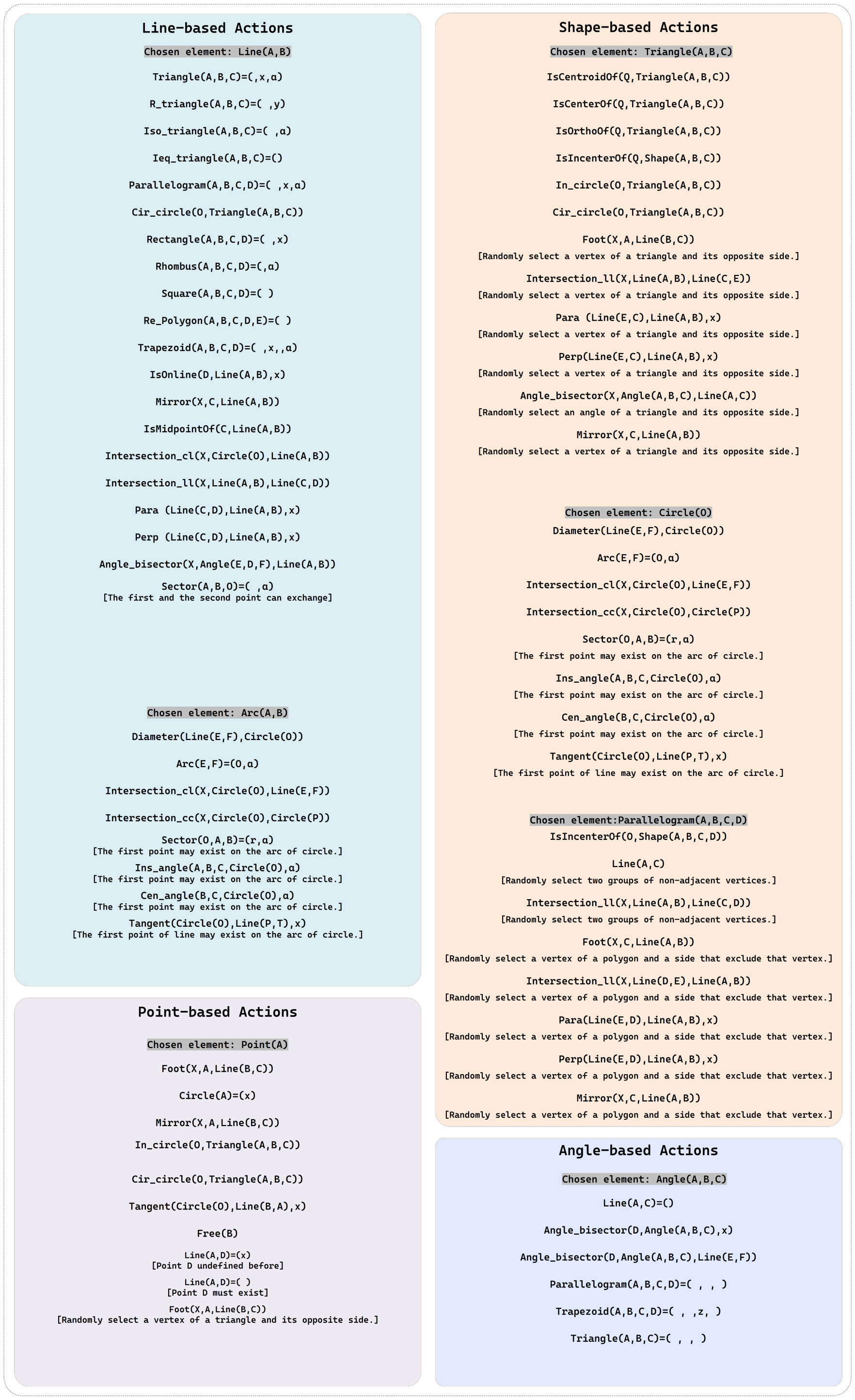
```
Use the mathematics you know to make simple and accurate inferences and get the conclusions based on image
descriptions.
Focus on extracting as much information as possible. Keep the final answer of all reasoning to one decimal
place.Use <conclusion> </conclusion> to include all your inferences.
Then, from your conclusions, select the three conclusions and answeers.Output them in the format <q1> </q1>
<a1> </a1> and so on.The answer to each question is only allowed to contain one number or angle.
Example is as follows:
Input:
Triangle SFD is a right triangle with the right angle at SFD.A right-angled triangle FSX with 90° angle at
FSX. Circle I is the incircle of triangle FSX. E is where the perpendicular from F meets line SD.
Output:
<conclusion>
1.The area of Triangle SFD is 18.0.
2.The length of line EF is 3.9.
3.The measure of angle DFS is 29.4°.
</conclusion>
<q1>What is the area of Triangle SFD?</q1> <a1>18.0</a1>
<q2>What is the length of line EF?</q2> <a2>3.9</a2>
<q3>What is the measure of angle DFS?</q3> <a3>29.4°</a3>
Here is the image description input: ‘{text-full}'
```

Figure 12. Prompt for Deepseek R1 in reverse search.

The prompt for DeepSeek V3 to forward validation

```
Here is a geometry problem with the following information:
Image description: ‘{text-full}’
Question: ‘{question from R1}’
Following the above condition and question,
think step by step and answer the following question directly.
After thinking process,you should provide your final concise reasoning steps in <steps></steps>
tags and your final answer in <answer></answer> tags.
The answer to the question is only allowed to contain one number or angle.
Keep the final answer of all reasoning to one decimal place.
Here is an example of your output:\n
Output:
<steps>
Step1: since AB is a segment from A to B, which would be AD + DB
AD is 3.5, and DB is x, so AB = 3.5 + x.
Step2:Since AB = DF (because the  triangles are congruent)
Step3:Let's call BF as y. Then DF = DB + BF = x + y.
we already said DF = AB = 3.5 + x. So according to that, x + y = 3.5 + x.
</steps>
<answer>
3.0
</answer>
```

Figure 13. Prompt for Deepseek V3 in forward validation.



Line-based Actions

```
Chosen element: Line(A,B)
Triangle(A,B,C)=(,x,α)
R_triangle(A,B,C)=( ,y)
Iso_triangle(A,B,C)=( ,α)
Ieq_triangle(A,B,C)=()
Parallelogram(A,B,C,D)=( ,x,α)
Cir_circle(O,Triangle(A,B,C))
Rectangle(A,B,C,D)=( ,x)
Rhombus(A,B,C,D)=(,α)
Square(A,B,C,D)=( )
Re_Polygon(A,B,C,D,E)=( )
Trapezoid(A,B,C,D)=( ,x,,α)
IsOnline(D,Line(A,B),x)
Mirror(X,C,Line(A,B))
IsMidpointOf(C,Line(A,B))
Intersection_cl(X,Circle(O),Line(A,B))
Intersection_ll(X,Line(A,B),Line(C,D))
Para (Line(C,D),Line(A,B),x)
Perp (Line(C,D),Line(A,B),x)
Angle_bisector(X,Angle(E,D,F),Line(A,B))
Sector(A,B,O)=( ,α)
[The first and the second point can exchange]

Chosen element: Arc(A,B)
Diameter(Line(E,F),Circle(O))
Arc(E,F)=(O,α)
Intersection_cl(X,Circle(O),Line(E,F))
Intersection_cc(X,Circle(O),Circle(P))
Sector(O,A,B)=(r,α)
[The first point may exist on the arc of circle.]
Ins_angle(A,B,C,Circle(O),α)
[The first point may exist on the arc of circle.]
Cen_angle(B,C,Circle(O),α)
[The first point may exist on the arc of circle.]
Tangent(Circle(O),Line(P,T),x)
[The first point of line may exist on the arc of circle.]
```

Point-based Actions

```
Chosen element: Point(A)
Foot(X,A,Line(B,C))
Circle(A)=(x)
Mirror(X,A,Line(B,C))
In_circle(O,Triangle(A,B,C))
Cir_circle(O,Triangle(A,B,C))
Tangent(Circle(O),Line(B,A),x)
Free(B)
Line(A,D)=(x)
[Point D undefined before]
Line(A,D)=( )
[Point D must exist]
Foot(X,A,Line(B,C))
[Randomly select a vertex of a triangle and its opposite side.]
```

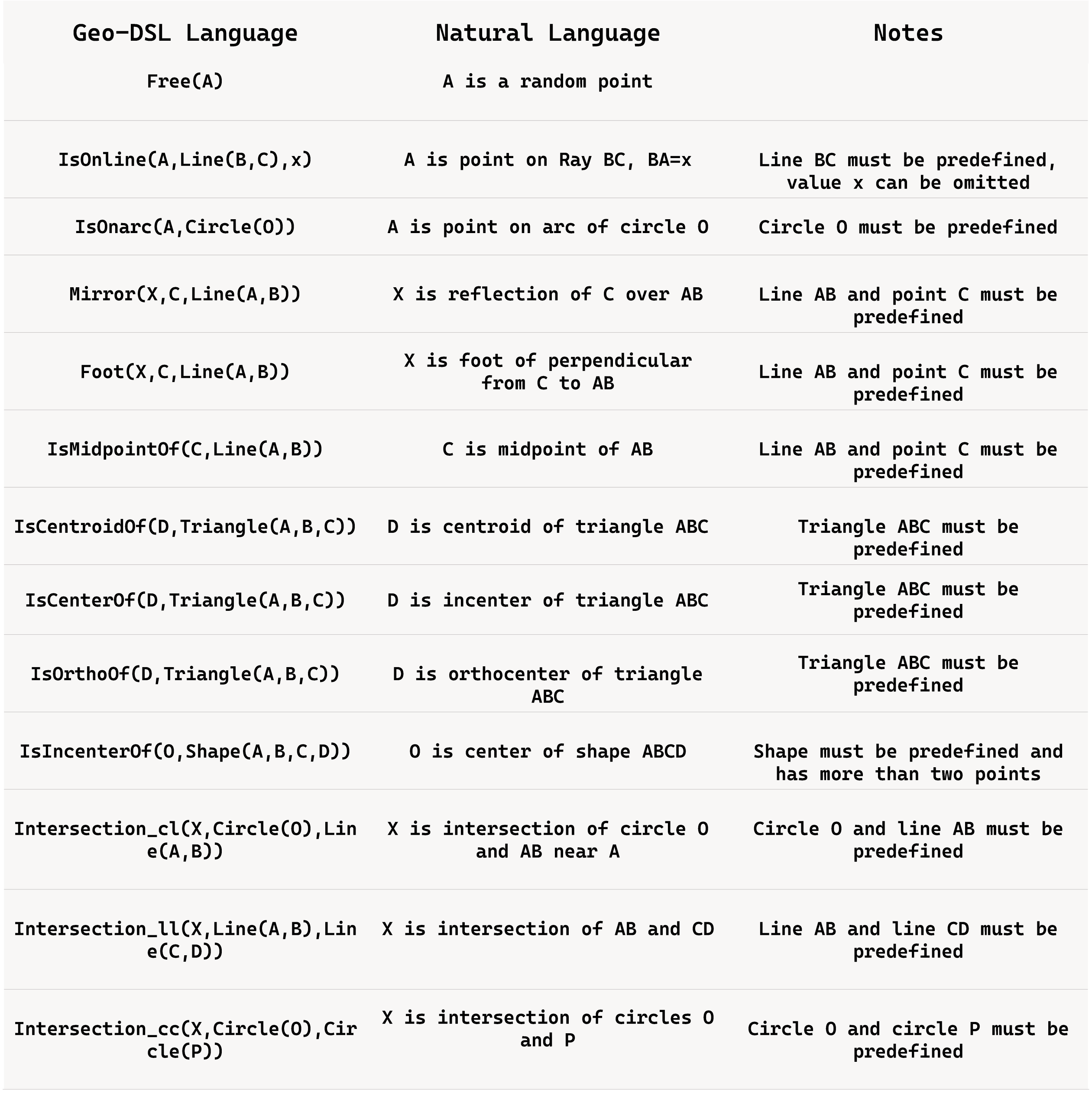
Shape-based Actions

```
Chosen element: Triangle(A,B,C)
IsCentroidOf(Q,Triangle(A,B,C))
IsCenterOf(Q,Triangle(A,B,C))
IsOrthoOf(Q,Triangle(A,B,C))
IsIncenterOf(Q,Shape(A,B,C))
In_circle(O,Triangle(A,B,C))
Cir_circle(O,Triangle(A,B,C))
Foot(X,A,Line(B,C))
[Randomly select a vertex of a triangle and its opposite side.]
Intersection_ll(X,Line(A,B),Line(C,E))
[Randomly select a vertex of a triangle and its opposite side.]
Para (Line(E,C),Line(A,B),x)
[Randomly select a vertex of a triangle and its opposite side.]
Perp(Line(E,C),Line(A,B),x)
[Randomly select a vertex of a triangle and its opposite side.]
Angle_bisector(X,Angle(A,B,C),Line(A,C))
[Randomly select an angle of a triangle and its opposite side.]
Mirror(X,C,Line(A,B))
[Randomly select a vertex of a triangle and its opposite side.]

Chosen element: Circle(O)
Diameter(Line(E,F),Circle(O))
Arc(E,F)=(O,α)
Intersection_cl(X,Circle(O),Line(E,F))
Intersection_cc(X,Circle(O),Circle(P))
Sector(O,A,B)=(r,α)
[The first point may exist on the arc of circle.]
Ins_angle(A,B,C,Circle(O),α)
[The first point may exist on the arc of circle.]
Cen_angle(B,C,Circle(O),α)
[The first point may exist on the arc of circle.]
Tangent(Circle(O),Line(P,T),x)
[The first point of line may exist on the arc of circle.]

Chosen element:Parallelogram(A,B,C,D)
IsIncenterOf(O,Shape(A,B,C,D))
Line(A,C)
[Randomly select two groups of non-adjacent vertices.]
Intersection_ll(X,Line(A,B),Line(C,D))
[Randomly select two groups of non-adjacent vertices.]
Foot(X,C,Line(A,B))
[Randomly select a vertex of a polygon and a side that exclude that vertex.]
Intersection_ll(X,Line(D,E),Line(A,B))
[Randomly select a vertex of a polygon and a side that exclude that vertex.]
Para(Line(E,D),Line(A,B),x)
[Randomly select a vertex of a polygon and a side that exclude that vertex.]
Perp(Line(E,D),Line(A,B),x)
[Randomly select a vertex of a polygon and a side that exclude that vertex.]
Mirror(X,C,Line(A,B))
[Randomly select a vertex of a polygon and a side that exclude that vertex.]
```

Angle-based Actions

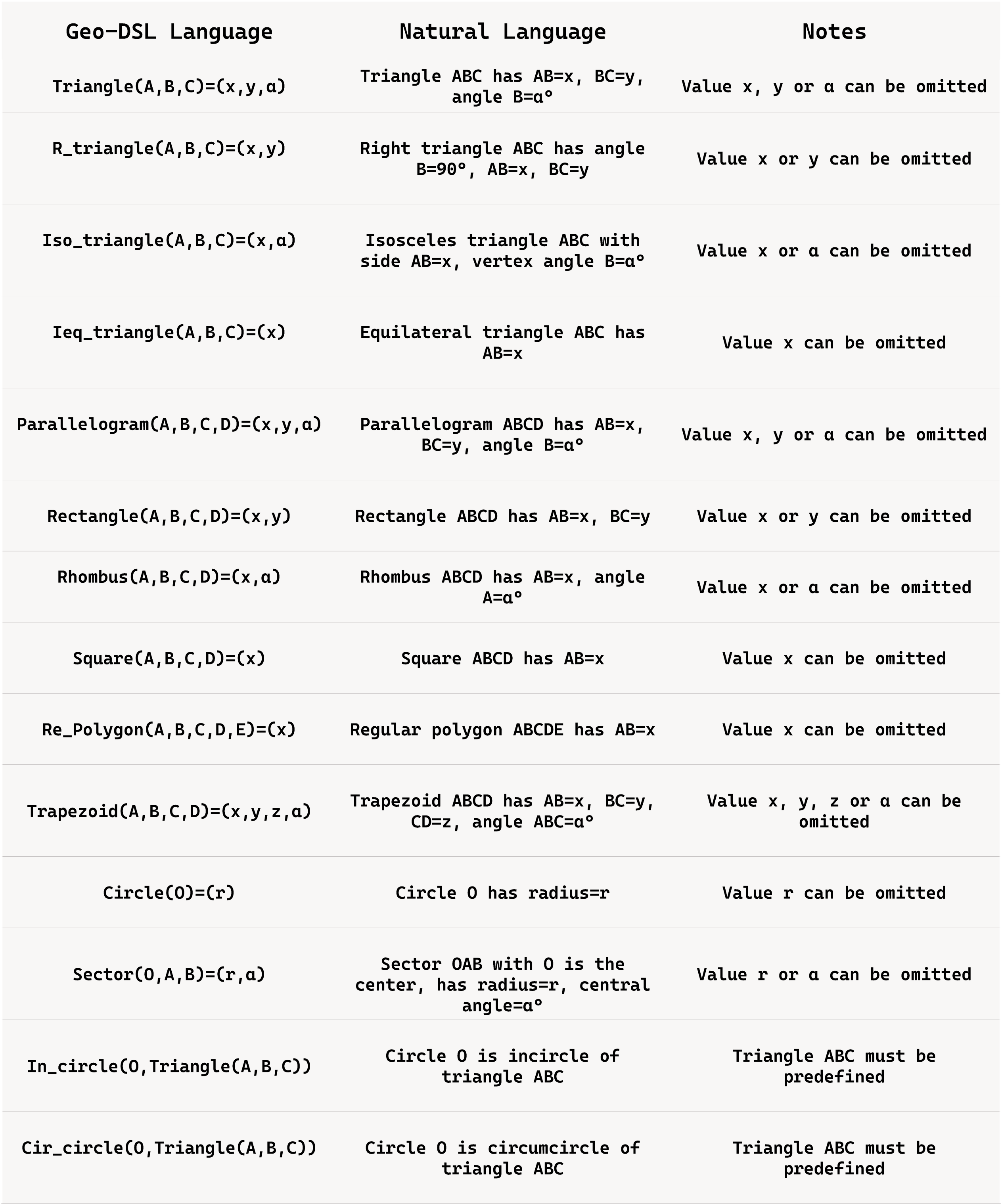
```
Chosen element: Angle(A,B,C)
Line(A,C)=()
Angle_bisector(D,Angle(A,B,C),x)
Angle_bisector(D,Angle(A,B,C),Line(E,F))
Parallelogram(A,B,C,D)=( , , )
Trapezoid(A,B,C,D)=( , ,z, )
Triangle(A,B,C)=( , , )
```

Figure 14. Detailed Actions in Symbolic Spaces. Actions can be categorized into four parts based on the type of selected geometric element: line-based, point-based, shape-based, and angle-based.



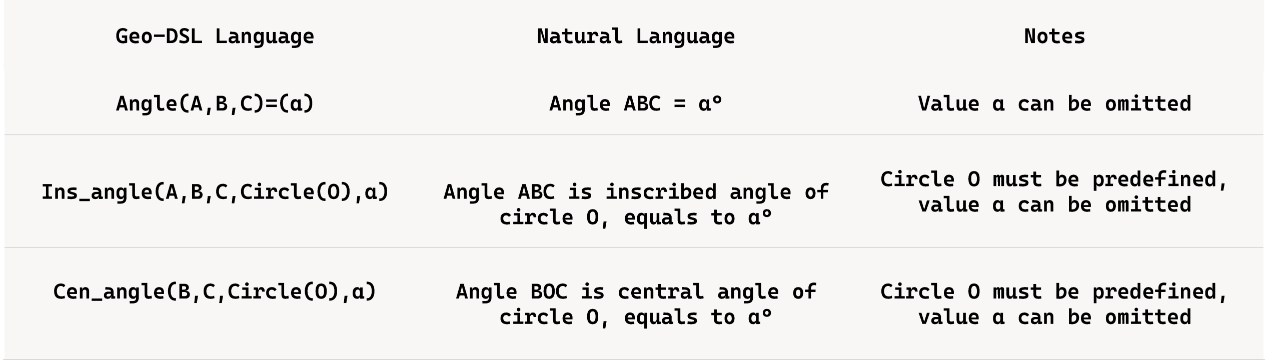
Geo-DSL Language	Natural Language	Notes
Free(A)	A is a random point	
IsOnline(A,Line(B,C),x)	A is point on Ray BC, BA=x	Line BC must be predefined, value x can be omitted
IsOnarc(A,Circle(O))	A is point on arc of circle O	Circle O must be predefined
Mirror(X,C,Line(A,B))	X is reflection of C over AB	Line AB and point C must be predefined
Foot(X,C,Line(A,B))	X is foot of perpendicular from C to AB	Line AB and point C must be predefined
IsMidpointOf(C,Line(A,B))	C is midpoint of AB	Line AB and point C must be predefined
IsCentroidOf(D,Triangle(A,B,C))	D is centroid of triangle ABC	Triangle ABC must be predefined
IsCenterOf(D,Triangle(A,B,C))	D is incenter of triangle ABC	Triangle ABC must be predefined
IsOrthoOf(D,Triangle(A,B,C))	D is orthocenter of triangle ABC	Triangle ABC must be predefined
IsIncenterOf(O,Shape(A,B,C,D))	O is center of shape ABCD	Shape must be predefined and has more than two points
Intersection_cl(X,Circle(O),Line(A,B))	X is intersection of circle O and AB near A	Circle O and line AB must be predefined
Intersection_ll(X,Line(A,B),Line(C,D))	X is intersection of AB and CD	Line AB and line CD must be predefined
Intersection_cc(X,Circle(O),Circle(P))	X is intersection of circles O and P	Circle O and circle P must be predefined

Figure 15. Geo-DSL definitions of line.



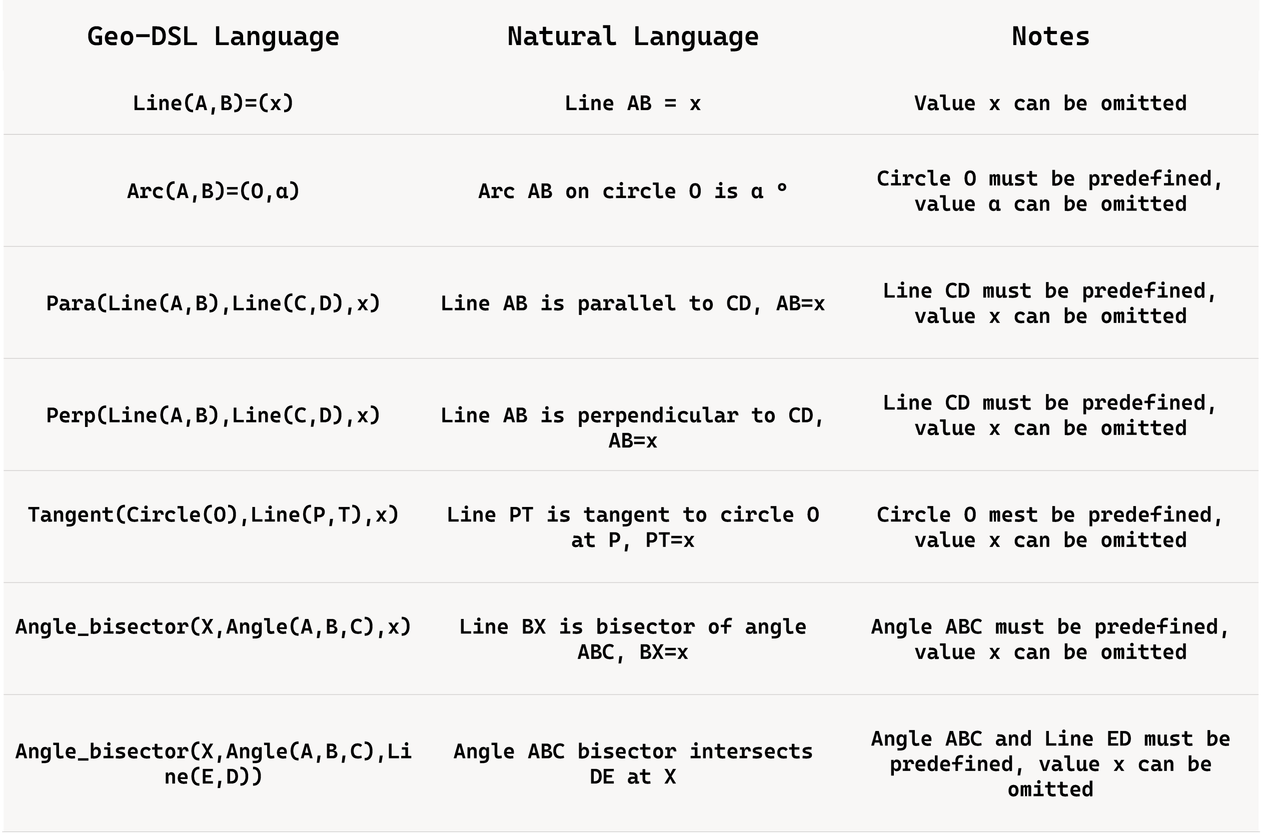
Geo-DSL Language	Natural Language	Notes
Triangle(A,B,C)=(x,y,α)	Triangle ABC has AB=x, BC=y, angle B=α°	Value x, y or α can be omitted
R_triangle(A,B,C)=(x,y)	Right triangle ABC has angle B=90°, AB=x, BC=y	Value x or y can be omitted
Iso_triangle(A,B,C)=(x,α)	Isosceles triangle ABC with side AB=x, vertex angle B=α°	Value x or α can be omitted
Ieq_triangle(A,B,C)=(x)	Equilateral triangle ABC has AB=x	Value x can be omitted
Parallelogram(A,B,C,D)=(x,y,α)	Parallelogram ABCD has AB=x, BC=y, angle B=α°	Value x, y or α can be omitted
Rectangle(A,B,C,D)=(x,y)	Rectangle ABCD has AB=x, BC=y	Value x or y can be omitted
Rhombus(A,B,C,D)=(x,α)	Rhombus ABCD has AB=x, angle A=α°	Value x or α can be omitted
Square(A,B,C,D)=(x)	Square ABCD has AB=x	Value x can be omitted
Re_Polygon(A,B,C,D,E)=(x)	Regular polygon ABCDE has AB=x	Value x can be omitted
Trapezoid(A,B,C,D)=(x,y,z,α)	Trapezoid ABCD has AB=x, BC=y, CD=z, angle ABC=α°	Value x, y, z or α can be omitted
Circle(O)=(r)	Circle O has radius=r	Value r can be omitted
Sector(O,A,B)=(r,α)	Sector OAB with O is the center, has radius=r, central angle=α°	Value r or α can be omitted
In_circle(O,Triangle(A,B,C))	Circle O is incircle of triangle ABC	Triangle ABC must be predefined
Cir_circle(O,Triangle(A,B,C))	Circle O is circumcircle of triangle ABC	Triangle ABC must be predefined

Figure 16. Geo-DSL definitions of shape.



Geo-DSL Language	Natural Language	Notes
Angle(A,B,C)=(α)	Angle ABC = α°	Value α can be omitted
Ins_angle(A,B,C,Circle(O),α)	Angle ABC is inscribed angle of circle O, equals to α°	Circle O must be predefined, value α can be omitted
Cen_angle(B,C,Circle(O),α)	Angle BOC is central angle of circle O, equals to α°	Circle O must be predefined, value α can be omitted

Figure 17. Geo-DSL definitions of angle.



Geo-DSL Language	Natural Language	Notes
Line(A,B)=(x)	Line AB = x	Value x can be omitted
Arc(A,B)=(O,α)	Arc AB on circle O is α °	Circle O must be predefined, value α can be omitted
Para(Line(A,B),Line(C,D),x)	Line AB is parallel to CD, AB=x	Line CD must be predefined, value x can be omitted
Perp(Line(A,B),Line(C,D),x)	Line AB is perpendicular to CD, AB=x	Line CD must be predefined, value x can be omitted
Tangent(Circle(O),Line(P,T),x)	Line PT is tangent to circle O at P, PT=x	Circle O mest be predefined, value x can be omitted
Angle_bisector(X,Angle(A,B,C),x)	Line BX is bisector of angle ABC, BX=x	Angle ABC must be predefined, value x can be omitted
Angle_bisector(X,Angle(A,B,C),Line(E,D))	Angle ABC bisector intersects DE at X	Angle ABC and Line ED must be predefined, value x can be omitted

Figure 18. Geo-DSL definitions of point.