

[Auxiliary Line Construction] Take A as the origin, and let the lines along AB, AD, and AF be the x-axis, y-axis, and z-axis, respectively, then establish the three-dimensional Cartesian coordinate system A-xyz.

Figure 5: Comparison of two representative image editing models for constructing a three-dimensional Cartesian coordinate system.

A THE USE OF LARGE LANGUAGE MODELS

In this paper, the authors used ChatGPT solely for language polishing, including grammar, phrasing, and stylistic refinement. We did not use it to generate scientific content, such as research ideas, methods, or related work. We did not provide any confidential, personal, or proprietary data to the model. The authors take full responsibility for all scientific content, which was exclusively written and verified by us.

B LIMITATION AND FUTURE WORK

Although our framework achieves competitive performance, it does not yet realize the ideal approach of directly rendering precise auxiliary lines on the diagram. Given the limited geometric controllability of current image editing and diffusion models, we instead employ a cross-modal reward model as a practical proxy to promote diagram-text consistency without directly editing the diagram images. In future work, we will investigate constraint-guided diffusion in conjunction with a geometry-constrained rendering engine to explicitly render auxiliary lines, thereby enhancing the alignment between visual constructions and symbolic reasoning.

C QUALITATIVE COMPARISON OF IMAGE EDITING MODELS FOR AUXILIARY-LINE GENERATION

To further highlight the limitations of current image-editing models in precise geometric construction, we present a single illustrative example comparing two representative models on a three-dimensional coordinate system construction task. As shown in Figure 5, this comparative example reveals a persistent difficulty in faithfully instantiating the specified auxiliary line descriptions, namely, aligning the edits with the intended spatial constraints, which motivates our text-driven auxiliary line construction guided by a cross-modal reward model.

D DATA CREATION

In this section, we detail the four progressive steps of our data creation pipeline.

D.1 AUTOMATED PROBLEM IDENTIFICATION

To construct the AuxSolidMath dataset, we first collect over 350 sets of high school geometry problems from publicly available online sources. Given that the dataset is intended to support constructive geometric reasoning, we specifically target problems that necessitate auxiliary line constructions as integral components of their solutions.

To efficiently identify such problems, we design an automated two-stage filtering pipeline using Python scripts. In the first stage, we detect problems whose solutions contain explicit mentions of auxiliary-line constructions. Specifically, we apply regular expression patterns to locate question number markers that are explicitly present in the exam papers and use these markers to segment the content into individual problem units. For each problem, we examine the solution for verbs that signal the introduction of auxiliary lines (e.g., "connect," "construct," "draw," "establish"). Problems lacking such terms are discarded, while those containing relevant cues are retained. In the second stage, we further refine the selection by ensuring that each retained problem contains both the original diagram and an auxiliary-line diagram. To this end, we quantify the number of diagrams associated with each problem. Problems with fewer than two diagrams are excluded, whereas those with at least two, which usually represent the original and modified diagrams, are preserved. This automated pipeline enables scalable and consistent filtering of auxiliary-line geometry problems, significantly reducing manual annotation effort.

715 D.2 AUTOMATED DEDUPLICATION AND DIAGRAM EXTRACTION

Upon identifying geometry problems requiring auxiliary lines, we employ an automated pipeline to deduplicate instances and extract the associated diagrams. This step guarantees the uniqueness and visual quality of data instances for downstream model training.

Problem Deduplication. To eliminate duplicate problems, we retain only the first occurrence of each unique problem based on its textual content. Concretely, we initialize a global problem set as an empty collection. We then sequentially process all Word-formatted exam papers, examining only the problem descriptions while ignoring the associated solutions and diagrams. For each problem, if its description is not already present in the global set, we add the problem; otherwise, we discard it as a duplicate. This procedure ensures that identical problems, which often recur across different examinations, are retained only once.

Diagram Extraction. Following deduplication, we extract, filter, and store the geometric diagrams associated with each retained problem. A key challenge lies in reliably distinguishing true geometric figures from image-embedded mathematical expressions (e.g., MathType equations), as both appear in Word exam papers. Existing Python libraries are unable to make this distinction accurately, often misclassifying equations as diagrams and introducing significant noise into the extraction process. To overcome this limitation, we innovatively integrate the Apache POI library through a custom Java implementation, enabling fine-grained control over the parsing of Word documents. This setup enables reliable identification and extraction of genuine geometric diagrams while effectively filtering out formula-rendered images. To further ensure visual quality, the extracted diagrams are then processed using OpenCV to discard low-resolution or unclear diagrams. The remaining diagrams are subsequently saved using a standardized naming convention that distinguishes between the original and the annotated versions of the auxiliary lines. To be more specific, for each problem indexed by i , we store two images: $\{i\}.png$, which contains the original diagram, and $\{i\}_\text{auxiliary}.png$, which includes the corresponding auxiliary-line diagram. This consistent format facilitates downstream alignment between textual and visual modalities within the multimodal processing pipeline.

742 D.3 STRUCTURED DATA EXTRACTION

Building on the high-quality geometric diagrams obtained in the previous step, we proceed to extract the corresponding textual content for each geometry instance, including the problem description, the final answer, and the auxiliary-line description. This extraction process is non-trivial, as the original Word documents frequently embed mathematical expressions using MathType formats that are not reliably handled by standard document parsing tools.

To address this challenge, we render the processed Word documents as images, thereby enabling LVLMs to leverage their visual reasoning capabilities. Although this approach appears straightforward, open-source models such as Qwen2.5-VL-7B-Instruct (Bai et al., 2025) often struggle to accurately parse complex geometry problems involving symbolic notation and mathematical expressions. In contrast, closed-source models like GPT-4o (Hurst et al., 2024) exhibit significantly stronger performance, but their reliance on commercial APIs introduces substantial costs and limits scalability in large-scale applications. To balance accuracy with scalability, we adopt a hybrid strategy. More concretely, we first utilize an advanced closed-source model (*i.e.*, GPT-4o) to gen-

756 erate a small, high-quality seed dataset comprising 300 manually verified instances. This curated
 757 dataset is then used to fine-tune an open-source LVLM (*i.e.*, Qwen2.5-VL-7B-Instruct), resulting
 758 in a lightweight, domain-adapted model capable of accurate and scalable text extraction. The final
 759 output consists of the extracted problem description, the final answer, and the auxiliary-line descrip-
 760 tion, all encapsulated in a structured JSON format. This unified representation facilitates consistent
 761 data handling and serves as a foundation for training a robust open-source text extraction model. By
 762 releasing this model, we aim to contribute a practical and reusable resource to the broader research
 763 community working on geometry-aware vision-language understanding.

764 D.4 MANUAL VERIFICATION

765 To ensure the quality and reliability of the final dataset, we perform a manual verification step that
 766 assesses each data instance in terms of accuracy, completeness, uniqueness, and semantic consis-
 767 tency, alongside visual quality criteria such as image clarity and resolution. Two AI researchers
 768 serve as independent checkers. Each instance is independently reviewed by both researchers. If
 769 either checker identifies a potential issue, the instance is collaboratively revised. This process is
 770 repeated up to three times per instance, ensuring that all errors, including subtle or ambiguous ones,
 771 are systematically identified and corrected. Manual verification plays a critical role in resolving
 772 complex symbolic expressions and ambiguous diagrammatic content that automated tools may mis-
 773 interpret. Despite its relatively low cost and effort, this step remains indispensable for ensuring the
 774 high data fidelity necessary for a reliable model.

775 E EXPERIMENTAL SETUP

776 E.1 MODELS

777 On the closed-source models, we include leading models such as gpt-5-mini (OpenAI, 2025),
 778 o4-mini (OpenAI, 2025) and GPT-4o (Hurst et al., 2024), Gemini-2.0-Flash and Gemini-
 779 2.5-Flash (DeepMind, 2025a), Claude 3.7 Sonnet (Anthropic, 2025a) and Claude Sonnet 4
 780 20250514 (Anthropic, 2025b). These models represent the forefront of multimodal reasoning among
 781 closed-source models, although their internal architectures remain undisclosed. On the open-source
 782 models, we consider several publicly available high-performance models, including the Qwen2
 783 VL (Wang et al., 2024b) and Qwen2.5 VL series (Bai et al., 2025), InternVL 3 families (Zhu et al.,
 784 2025), LLaMA-3.2-11B-Vision-Instruct (Meta, 2024) and Llama-4-Scout-17B-16E-Instruct (Meta,
 785 2025). These models encompass a range of design paradigms, parameter scales, and instruction
 786 tuning strategies, providing a robust comparative foundation for evaluating multimodal reasoning
 787 capabilities. Note that models such as Gemini-2.5 Pro (DeepMind, 2025b) and OpenAI o3 (Ope-
 788 nAI, 2025) are excluded from our study due to limited accessibility and high inference costs.

789 E.2 TRAINING IMPLEMENTATION DETAILS

790 We adopt a two-stage training paradigm based on the Qwen2.5-VL series, including Qwen2.5-VL-
 791 3B and Qwen2.5-VL-7B, consisting of the SFT stage and the RL stage.

792 **SFT Stage.** The SFT phase is conducted using the LLaMA-Factory framework (Zheng et al., 2024).
 793 For Qwen2.5-VL-7B, we train the model for 5 epochs with a per-device batch size of 2 and a gradient
 794 accumulation step of 8 (effective batch size of 16). We use the AdamW optimizer with a learning
 795 rate of 2e-5 and apply a cosine learning rate scheduler with a warmup ratio of 0.1. The model is
 796 trained in bf16 precision. Vision and projection modules are frozen during this stage, while the
 797 language model remains unfrozen. For Qwen2.5-VL-3B, we adopt the same training configuration
 798 as the 7B variant, except learning rate and training epochs. Specifically, Qwen2.5-VL-3B is trained
 799 for 5 epochs with a learning rate of 3e-5.

800 **RL Stage.** The RL phase is performed using the EasyR1 framework (Zheng et al., 2025) with the
 801 GRPO algorithm. For the Qwen2.5-VL-7B model, training and validation data are loaded from
 802 Parquet files containing question-diagram pairs, with a maximum response length of 8192. Both
 803 rollout and validation batch size are set to 16. The actor is optimized using AdamW (learning
 804 rate 2e-6, weight decay 1e-2, no warmup). KL regularization is applied using the `low_var_kl`

810 penalty with a coefficient of 1e-2. Training runs for 6 epochs using bf16 precision, with gradient
 811 checkpointing and partial FSDP offloading enabled for memory efficiency.
 812

813 *Rewards.* The overall reward is the sum of a cross-modal auxiliary-line consistency reward and
 814 a final-answer accuracy reward, where the auxiliary-line component is weighted by $\alpha = 0.1$. For
 815 Qwen2.5-VL-3B, we adopt the same RL configuration as the 7B model, with adjustments to the
 816 batch size and the number of training epochs. Specifically, Qwen2.5-VL-3B is trained for 4 epochs
 817 with a batch size of 8.

818 All training was conducted on a server equipped with two NVIDIA A100 80GB and two NVIDIA
 819 A800 80GB GPUs. The SFT stage was performed on the A100 GPUs, while the full set of four
 820 GPUs was utilized during the reinforcement learning stage.

822 F RELATED WORK

823 In this section, we also review benchmarks and datasets for geometric reasoning. Most benchmarks
 824 and datasets for geometric problem solving focus on plane geometry, where diagrams and prob-
 825 lems involve two-dimensional figures. Well-known resources in this area include Geometry3K (Lu
 826 et al., 2021), GeoQA (Chen et al., 2021a), UniGeo (Chen et al., 2022), and GeomRel (Wang et al.,
 827 2025b), which primarily cover plane geometry problems. A concurrent benchmark, GeoLaux (Fu
 828 et al., 2025), explores the use of auxiliary lines in plane geometry, but is limited to simple cases
 829 and lacks engagement with the spatial complexity of solid geometry. Nonetheless, there remains a
 830 lack of dedicated resources for solid geometry, even though solving such problems often requires
 831 interpreting three-dimensional relationships and drawing auxiliary lines to uncover hidden spatial
 832 structures. While SolidGeo (Wang et al., 2025a) is a recent benchmark that focuses exclusively on
 833 solid geometry, it does not explicitly require auxiliary lines for solving its problems, leaving this
 834 important aspect of spatial reasoning underexplored. Similarly, other benchmarks such as Math-
 835 Vista (Lu et al., 2024), MathVision(Wang et al., 2024a), and MathVerse (Zhang et al., 2024) contain
 836 only a limited number of solid geometry problems, and these also do not require auxiliary lines to
 837 reach the solution. As a result, these resources fall short of evaluating a model’s ability to solve
 838 complex solid geometry problems where auxiliary lines are essential for uncovering implicit spatial
 839 relationships. To address this gap, we present AuxSolidMath, the first dedicated dataset for solid
 840 geometry problems that require auxiliary lines to solve. It offers comprehensive multimodal super-
 841 vision, including the original diagram, the problem statement, textual descriptions of the required
 842 auxiliary lines, the final answer, and a corresponding diagram annotated with those lines, enabling
 843 models to learn how the auxiliary lines facilitate solid geometry reasoning.

845 G PROMPTS

846 G.1 PROMPTS FOR SUPERVISED FINE-TUNING

847 The following presents the two-part prompt template used in our supervised dataset. The sys-
 848 tem prompt assigns the solver role and enforces formatting: auxiliary lines must be wrapped in
 849 `[AUX] ... [/AUX]` and the final answer must appear as plain text in `Final Answer:....`. The user
 850 prompt is multimodal, pairing a diagram referenced by the `<image>` token with the natural language
 851 question `{question}`, which yields explicit reasoning steps and a final answer.

852 SYSTEM_PROMPT_FOR_SFT

853 `SYSTEM_PROMPT_FOR_SFT = """`

854 You are a mathematician skilled in solving geometry problems through step-by-step reasoning. Solve
 855 the given geometry problem based on a geometric diagram and a natural language question. Use
 856 `'[AUX]...[/AUX]'` to indicate auxiliary constructions, such as establishing coordinate systems or con-
 857 structing auxiliary lines. Finally, provide your final answer within `'Final Answer:....'`.
 858
 859 `"""`

```

864 USER_PROMPT_FOR_SFT
865
866 USER_PROMPT_FOR_SFT = """
867 Image: <image>
868 Question: {question}
869 """
870

```

871 G.2 PROMPTS FOR CROSS-MODAL REWARD MODEL

873 Using the prompt below, the cross-modal reward model compares the description of the auxiliary line
 874 generated by the policy model against a pair of diagrams, the original image I and its auxiliary-line
 875 counterpart I^+ , and returns a single line justification and a calibrated score in $[0, 1]$ that measures
 876 visual-textual agreement. The instruction emphasizes the correctness of auxiliary-line constructions
 877 and adherence to geometric constraints. Higher scores indicate stronger alignment.

879 **SYSTEM_PROMPT_FOR_CROSS-MODAL_REWARD_MODEL**

```

880
881 SYSTEM_PROMPT_FOR_CROSS-MODAL_REWARD_MODEL = """
882 You are a professional geometry reasoning evaluator. Your task is to evaluate whether a given textual
883 description of auxiliary lines accurately explains the visual difference between the original diagram and
884 the auxiliary-line diagram.
885
886 Score the description on a scale from 0 to 1:
887 1. 1 indicates a fully accurate and helpful description.
888 2. 0 indicates a completely irrelevant or misleading description.
889 3. Intermediate values (e.g., 0.25/ 0.50/ 0.75) reflect partial relevance or minor issues.
890
891 Return exactly one line:
892 <brief justification>. Score: <s>.
893 """
894

```

894 **USER_PROMPT_FOR_CROSS-MODAL_REWARD_MODEL**

```

895
896 USER_PROMPT_FOR_CROSS-MODAL_REWARD_MODEL = """
897 Image (original diagram): <image  $I$ >
898 Image (auxiliary-line diagram): <image  $I^+$ >
899 Auxiliary-line description: {generated_aux_description}
900 """
901

```

902 H REPRESENTATIVE EXAMPLES

904 As illustrad in Figure 6 present qualitative examples from the AuxSolidMath dataset, including the
 905 question, the final answer, the auxiliary-line description, the original diagram, and the auxiliary-line
 906 diagram. The examples showcase diverse strategies for constructing auxiliary lines and demonstrate
 907 that explicit annotations reveal the key spatial constraints.

```

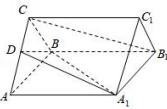
908
909
910
911
912
913
914
915
916
917

```


972 I CASE STUDY
973974 I.1 SUCCESSFUL CASES
975

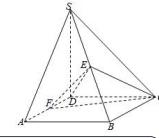
976 In this section, we present representative cases produced by GeoVLM-7B (see Figure 7). The
977 examples illustrate how valid auxiliary-line constructions capture essential spatial constraints, which
978 in turn lead to correct final answers. Collectively, these results provide concrete evidence of both
979 the effectiveness of our training framework and the model’s capacity to tackle complex geometric
980 reasoning requiring robust visual understanding and precise symbolic inference.

981
982 **Question:** As shown in the diagram, in the regular triangular prism $ABC - A_1 B_1 C_1$, the side length of
983 the base is 2, and the lateral edge length is $\sqrt{3}$. D is the midpoint of AC. Find the volume of the triangular
984 pyramid $D - A_1 C_1 B$.



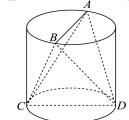
985
986 **Solution:** ∇ [AUX]1. **Construct auxiliary lines:** Take the midpoint E of BC_1 , and connect $A_1 E$ and $D_1 E$. ∇ [AUX] ∇ 2.
987 Determine parallel lines: Since $ABC-A_1 B_1 C_1$ is a regular triangular prism, $AB \parallel C_1 D_1$ and $AB = C_1 D_1$. So
988 quadrilateral $ABC_1 D_1$ is a parallelogram. ∇ 3. Identify equal lengths: Given the side length of the base is 2 and the lateral edge
989 length is $\sqrt{3}$, triangle $C_1 DC$ is equilateral with side length 2. ∇ 4. Calculate the volume: Since $ABC-A_1 B_1 C_1$ is a regular
990 triangular prism and triangle ABC is equilateral with side length 2, the height of the pyramid from D to the plane $A_1 C_1 B$ can be
991 calculated as $\sqrt{3}$. The area of the triangular base $A_1 C_1 B$ is $(\sqrt{3})/4 * (2)^2 = \sqrt{3}$. Therefore, the volume of the pyramid is $(1/3) *$
992 base area * height = $(1/3) * \sqrt{3} * \sqrt{3} = 1$. ∇ **Final Answer:** 1.

993
994 **Question:** As shown in the diagram, in the quadrilateral pyramid S - ABCD, SD is perpendicular to the
995 plane ABCD. The base ABCD is a rhombus, with E and F being the midpoints of SB and AD, respectively.
996 If $\angle BAD = 60^\circ$, $SD = 4$, and $AB = 2$, find the volume of the triangular pyramid C - DEF.



997
998 **Solution:** ∇ [AUX]1. **Construct auxiliary lines:** Take the midpoint H of AB, and connect DH. ∇ 2. Since $\angle BAD = 60^\circ$ and $AB = AD = 2$, triangle ABD is equilateral. ∇ 3. Since H is the midpoint of AB, $AH = 1$. By Pythagoras theorem in right
1000 triangle ADH, $DH = \sqrt{(AD^2 - AH^2)} = \sqrt{(2^2 - 1^2)} = \sqrt{3}$. ∇ 4. Since ABCD is a rhombus with $AB = BC = CD = DA = 2$ and
1001 $\angle ADC = 120^\circ$, we can calculate DC. Using cosine rule in triangle ADC, $DC^2 = AD^2 + AC^2 - 2*AD*AC*\cos(120^\circ) = 2^2 + (2^2) - 2*2*2*\cos(\pi/3) = 4 + 4 - 4 = 1$. So $DC = 1$. ∇ 5. Triangle CDE has vertices: C, D, E. E is the midpoint of SB. Since S is
1002 the apex and SD \perp plane ABCD, plane SAB \perp plane ABCD. Therefore, DE \perp plane ABCD. Since CH lies in plane ABCD, we
1003 conclude DE \perp CH. ∇ 6. Since AB = BC = CD = DA = 2, and H is the midpoint of AB, AC \perp BH. Similarly, since CD \perp DA
1004 (because $AD^2 + CD^2 = 2^2 + 1^2 = 5$, $AC^2 = 2^2 - 1^2 = 3$) and DC = 1, we can deduce CH \perp CD. Thus CH \perp plane
1005 CDE. ∇ 7. The area of triangle CDE is: ∇ Area(CDE) = $(1/2) * DC * DE = (1/2) * 1 * 2 = 1$. ∇ 8. The height of the pyramid C-
1006 DEF is CH. We need to find CH. In triangle CDB, DB = $2\sqrt{2}$ (by Pythagoras theorem). So $CH = AC \sin(60^\circ) = 2 * (\sqrt{3}/2) = \sqrt{3}$. ∇ 9. The volume of pyramid C-DEF is: ∇ Volume(C-DEF) = $(1/3) * \text{Area}(CDE) * CH = (1/3) * 1 * \sqrt{3} = \sqrt{3}/3$. ∇
1007 **Final Answer:** $\sqrt{3}/3$.

1008
1009 **Question:** As shown in the diagram, the diameter of the base of the cylinder is 3, and the slant height is 4.
1010 AB and CD are the diameters of the upper and lower bases of the cylinder, respectively, and $AB \perp CD$.
1011 Find the volume of the triangular pyramid A-BCD.



1012
1013 **Solution:** ∇ [AUX]1. **Construct auxiliary lines:** Take the midpoint M of AB, and connect MD and MC. ∇ 2.
1014 Determine the base area of the triangular pyramid: Since $AB \perp CD$ and $AB = CD = 3$ (diameter), the area of the triangular base
1015 ABCD is given by: ∇ Area = $(1/2) * AB * CD = (1/2) * 3 * 3 = 4.5$. ∇ 3. Determine the height of the pyramid: The height of the
1016 triangular pyramid A-BCD is the perpendicular distance from A to the plane BDC. Given that $AB \perp$ plane BDC, the height is AB
1017 itself, which is given as 4 (slant height). ∇ 4. Calculate the volume of the triangular pyramid A-BCD: ∇ Volume = $(1/3) * \text{Area}$
1018 of base * Height = $(1/3) * 4.5 * 4 = 6$. ∇ **Final Answer:** 6.

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

1065

1066

1067

1068

1069

1070

1071

1072

1073

1074

1075

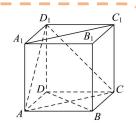
1076

1077

1078

1079

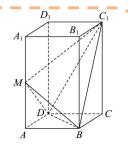
Question: As shown in the diagram, the edge length of the cube ABCD – A₁ B₁ C₁ D₁ is 2. Find the sine of the angle between line BD and plane ACD₁.



Solution: **[AUX]1. Establish coordinate system: Take point D as the origin, and let the lines DA, DC, and DD₁ be the x-axis, y-axis, and z-axis, respectively, to establish the three-dimensional Cartesian coordinate system D-xyz.** **[/AUX]**

Determine coordinates of points: Since the edge length of the cube ABCD – A₁ B₁ C₁ D₁ is 2, we have the coordinates of the vertices as follows: D = (0, 0, 0) A = (2, 0, 0) C = (0, 2, 0) B = (2, 2, 0) D₁ = (0, 0, 2) A₁ = (2, 0, 2) C₁ = (0, 2, 2) B₁ = (2, 2, 2). Find the coordinates of point E: E is the midpoint of BD. The coordinates of B are (2, 2, 0) and D are (0, 0, 0). The midpoint E is the average of the coordinates of B and D: E = ((2+0)/2, (2+0)/2, (0+0)/2) = (1, 1, 0). Find the vector BD: We have point B = (2, 2, 0) and point D = (0, 0, 0). Vector BD = D - B = (0-2, 0-2, 0-0) = (-2, -2, 0). Find a normal vector to the plane ACD₁: To find the normal vector, we need two vectors in the plane ACD₁. Vector AC = C - A = (0-2, 2-0, 0-0) = (-2, 2, 0). Vector AD₁ = D₁ - A = (0-2, 0-0, 2-0) = (-2, 0, 2). Normal vector n to the plane ACD₁ is the cross product of AC and AD₁: n = AC x AD₁ = (-2, 2, 0) x (-2, 0, 2) = (2*2 - 0*0, 0*(-2) - (-2)*2, (-2)*0 - 2*(-2)) = (4, 4, 4). We can take a simpler normal vector by dividing by 4, so we can take n = (1, 1, 1). Calculate the sine of the angle between line BD and plane ACD₁: The sine of the angle θ between line BD and plane ACD₁ is given by the absolute value of the cosine of the angle between vector BD and normal vector n. $\sin(\theta) = |\cos(\text{angle between BD and n})| = |(BD \cdot n) / (|BD| * |n|)|$. BD · n = (-2, -2, 0) · (1, 1, 1) = (-2)(1) + (-2)(1) + (0)(1) = -2 - 2 + 0 = -4. |BD| = sqrt((-2)^2 + (-2)^2 + 0^2) = sqrt(4 + 4) = sqrt(8) = 2sqrt(2). |n| = sqrt(1^2 + 1^2 + 1^2) = sqrt(3). sin(theta) = |-4| / (2sqrt(2) * sqrt(3)) = |-4| / (2sqrt(6)) = |-2sqrt(6)| / 2sqrt(6) = 2sqrt(6) / 6 = sqrt(6) / 3. **Final Answer:** **sqrt(6) / 3.**

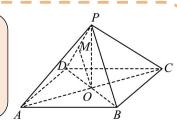
Question: As shown in the diagram, in the regular quadrilateral prism ABCD-A₁ B₁ C₁ D₁, M is the midpoint of AA₁, AB = 2, AA₁ = 4. Find the cosine value of the dihedral angle C₁ - BD - M.



Solution: **[AUX]1. Establish coordinate system: Take point D as the origin, and let the lines DA, DC, and DD₁ be the x-axis, y-axis, and z-axis, respectively, to establish the three-dimensional Cartesian coordinate system D-xyz.** **[/AUX]**

Determine coordinates of points: Since ABCD-A₁ B₁ C₁ D₁ is a regular quadrilateral prism with base side length AB = 2 and height AA₁ = 4, we have the coordinates of the vertices as follows: D = (0, 0, 0) A = (2, 0, 0) C = (0, 2, 0) B = (2, 2, 0) D₁ = (0, 0, 4) A₁ = (2, 0, 4) M is the midpoint of AA₁, so M = ((2+2)/2, (0+0)/2, (0+4)/2) = (2, 0, 2). Find normal vectors of planes C₁ - BD and MBD: For plane C₁ - BD, we need two vectors in the plane: C₁ - D and C₁ - B. Vector C₁ - D = D - C₁ = (0, 0, 0) - (0, 2, 4) = (0, -2, -4). Vector C₁ - B = B - C₁ = (2, 2, 0) - (0, 2, 4) = (2, 0, -4). The normal vector n1 of plane C₁ - BD is the cross product of C₁ - D and C₁ - B: n1 = C₁ - D x C₁ - B = (0, -2, -4) x (2, 0, -4) = (-2)(-4) - (-4)(0) - (0)(-4) = (8, -8, 4). We can take a simpler normal vector by dividing by 4, so n1 = (2, -2, 1). For plane MBD, we need two vectors in the plane: MB and MD. Vector MB = B - M = (2, 2, 0) - (2, 0, 2) = (0, 2, -2). Vector MD = D - M = (0, 0, 0) - (2, 0, 2) = (0, 2, -2). The normal vector n2 of plane MBD is the cross product of MB and MD: n2 = MB x MD = (2, 2, -2) x (0, 2, -2) = (2(-2) - (-2)(-2), (-2)(-2) - (0)(-2), (0)(-2) - (-2)(-2)) = (-4, 4, 4). We can take a simpler normal vector by dividing by 4, so n2 = (-1, 1, 1). Calculate the cosine of the dihedral angle: The cosine of the dihedral angle θ is the cosine of the angle between the normal vectors n1 and n2. cos(theta) = |(n1 · n2) / (|n1| * |n2|)|. n1 · n2 = (2, -2, 1) · (-1, 1, 1) = -2 - 2 + 1 = -3. |n1| = sqrt(2^2 + (-2)^2 + 1^2) = sqrt(9) = 3. |n2| = sqrt((-1)^2 + 1^2 + 1^2) = sqrt(3). cos(theta) = |-3| / (3 * sqrt(3)) = |-3| / (3 * sqrt(3)) = |-1| / sqrt(3) = 1 / sqrt(3) = sqrt(3) / 3. **Final Answer:** **sqrt(3) / 3.**

Question: In the quadrilateral pyramid P-ABCD, the quadrilateral ABCD is a rhombus, with AC and BD intersecting at point O. PA = PC, PB = PD, $\angle BAD = 60^\circ$, AB = 2, and M is the midpoint of segment PD. If the angle between line OM and plane ABCD is 60° , find the sine of the dihedral angle between plane PAD and plane PBC.



Solution: **[AUX]1. Establish coordinate system: Take point O as the origin, and let the lines OA, OB, and OP be the x-axis, y-axis, and z-axis, respectively, to establish the three-dimensional Cartesian coordinate system O-xyz.** **[/AUX]**

Determine coordinates of points: Since the rhombus ABCD has side length 2 and $\angle BAD = 60^\circ$, triangle ABD is equilateral, and triangle BCD is also equilateral. The diagonals AC and BD are perpendicular bisectors of each other. Given AB = 2, we have AO = OC = 1 and BO = OD = sqrt(2^2 - 1^2) = sqrt(3). Thus, the coordinates are: O = (0, 0, 0) A = (1, 0, 0) C = (-1, 0, 0) B = (0, sqrt(3), 0). Since PA = PC and PA = sqrt((1-0)^2 + (0-0)^2 + (sqrt(3)-0)^2) = sqrt(10), PC = sqrt((-1-0)^2 + (0-0)^2 + (sqrt(3)-0)^2) = sqrt(10). Squaring both PA and PC and equating them: (1-h)^2 + k^2 = (-1-h)^2 + k^2. Substituting h = 0 into PA^2 = (1-0)^2 + (0-0)^2 + k^2 = 1 + k^2 = 1^2 = 1 + (h^2) = PB^2. So z coordinate of P = OM = sin(60 degrees) = sqrt(3)/2. M is the midpoint of PD, M = ((0+0)/2, (0+0)/2, sqrt(3)/2) = (0, 0, sqrt(3)/2). P = (0, 0, sqrt(3)). A = (1, 0, 0), B = (0, sqrt(3), 0), C = (-1, 0, 0), D = (0, -sqrt(3), 0), P = (0, 0, sqrt(3)). Find normal vectors of plane PAD and plane PBC: For plane PAD, we need vectors PA and PD. PA = A - P = (1, 0, -sqrt(3)). PD = D - P = (0, -sqrt(3), -sqrt(3)). Normal vector n1 of plane PAD is the cross product of PA and PD: n1 = PA x PD = (1, 0, -sqrt(3)) x (0, -sqrt(3), -sqrt(3)) = (-sqrt(3), -sqrt(3), sqrt(3)). For plane PBC, we need vectors PB and PC. PB = B - P = (0, sqrt(3), -sqrt(3)). PC = C - P = (-1, 0, -sqrt(3)). Normal vector n2 of plane PBC is the cross product of PB and PC: n2 = PB x PC = (0, sqrt(3), -sqrt(3)) x (-1, 0, -sqrt(3)) = (sqrt(3), sqrt(3), sqrt(3)). Calculate the cosine of the dihedral angle: cos(theta) = |(n1 · n2) / (|n1| * |n2|)|. n1 · n2 = (-sqrt(3), -sqrt(3), sqrt(3)) · (sqrt(3), sqrt(3), sqrt(3)) = -3. |n1| = sqrt(3^2 + 3^2 + 3^2) = sqrt(27) = 3sqrt(3). |n2| = sqrt(3^2 + 3^2 + 3^2) = sqrt(27) = 3sqrt(3). cos(theta) = |-3| / (3sqrt(3) * 3sqrt(3)) = |-3| / (27) = -1/9. Calculate the sine of the dihedral angle: sin^2(theta) = 1 - cos^2(theta) = 1 - (-1/9)^2 = 1 - 1/81 = 80/81. sin(theta) = sqrt(80/81) = 4sqrt(5)/9. **Final Answer:** **4sqrt(5)/9.**

Figure 7: Successful cases generated by GeoVLMath-7B.

1080
1081

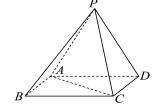
I.2 FAILURE CASES

1082
1083
1084
1085
1086
1087

In this section, we illustrate typical failure cases generated by GeoVLMath-7B (see Figure 8). The observed failures are exemplified by mis-specified coordinate systems and auxiliary-line descriptions that are irrelevant to the diagram, both of which fail to capture essential spatial constraints and may yield incorrect final answers. To mitigate such errors, we will explore diffusion-based drawing modules that render auxiliary lines directly on the original diagram and support iterative correction as part of future work.

1088
1089
1090
1091
1092
1093
1094

Question: As shown in the diagram, in the quadrilateral pyramid P-ABCD, the base ABCD is a square, PAD is an equilateral triangle, and $PC = AC$. Find the sine of the angle between line PB and plane PCD.



Ground Truth: $\sqrt{6}/4$

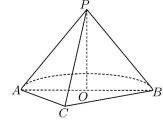
Analysis: The A-xyz coordinate system is incorrectly established—**AP is not orthogonal to AB or AD**, violating the pairwise orthogonality required for a three-dimensional Cartesian system. Consequently, the coordinates assigned to points and lines are invalid, leading to an incorrect final answer.

1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106

Solution: **[AUX]1. Establish coordinate system: Take point A as the origin, and let the lines AB, AD, and AP be the x-axis, y-axis, and z-axis, respectively, to establish the three-dimensional Cartesian coordinate system A-xyz.** **[/AUX]** $\text{Determine the coordinates of the vertices: Let the side length of the square ABCD be 2. Since A is the origin, } A = (0, 0, 0). \text{ As AB is along the x-axis and B is to the right of A with length 2, } B = (2, 0, 0). \text{ As AD is along the y-axis and D is above A with length 2, } D = (0, 2, 0). \text{ Since ABCD is a square, } C = (2, 2, 0). \text{ As PA is a line from A in the z-direction upwards and triangle PAD is an equilateral triangle with side length 2, the height from P to the base ABCD must be } \sqrt{(2^2 - 1^2)2} = \sqrt{3}. \text{ So, } P = (0, 0, \sqrt{3}).$ $\text{Find the vector PB: We have } P = (0, 0, \sqrt{3}) \text{ and } B = (2, 0, 0). \text{ The vector } PB = B - P = (2, 0, 0) - (0, 0, \sqrt{3}) = (2, 0, -\sqrt{3}).$ $\text{Find the normal vector of the plane PCD: We need two vectors in the plane PCD, for example, vector PC and vector PD. Vector } PC = C - P = (2, 2, 0) - (0, 0, \sqrt{3}) = (2, 2, -\sqrt{3}). \text{ Vector } PD = D - P = (0, 2, 0) - (0, 0, \sqrt{3}) = (0, 2, -\sqrt{3}). \text{ The normal vector } n \text{ of the plane PCD is the cross product of } PC \text{ and } PD: n = PC \times PD = (2, 2, -\sqrt{3}) \times (0, 2, -\sqrt{3}) = (2 * (-\sqrt{3}) - (-\sqrt{3}) * 2, (-\sqrt{3}) * 0 - 2 * (-\sqrt{3}), 2 * 2 - 2 * 0) = (-2\sqrt{3} + 2\sqrt{3}, 0 + 2\sqrt{3}, 4 - 0) = (0, 2\sqrt{3}, 4).$ $\text{To take a simpler normal vector, we can use } n' = (0, \sqrt{3}, 2).$ $\text{Calculate the sine of the angle between line PB and plane PCD: The sine of the angle } \theta \text{ between line PB and plane PCD is given by the absolute value of the cosine of the angle between vector PB and normal vector } n': \text{nsin}(0) = |\cos(\text{angle between } PB \text{ and } n')| = |(PB \cdot n') / (|PB| * |n'|)|.$ $\text{Vector } PB \cdot n' = (2, 0, -\sqrt{3}) \cdot (0, \sqrt{3}, 2) = 2 * 0 + 0 * \sqrt{3} + (-\sqrt{3}) * 2 = -2\sqrt{3}. |PB| = \sqrt{(2^2 + 0^2 + (-\sqrt{3})^2)} = \sqrt{(4 + 0 + 3)} = \sqrt{7}. |n'| = \sqrt{(0^2 + (\sqrt{3})^2 + 2^2)} = \sqrt{(0 + 3 + 4)} = \sqrt{7}.$ $\text{nsin}(0) = | -2\sqrt{3} / \sqrt{7} | = | -2\sqrt{3} / 7 | = 2\sqrt{3} / 7.$ **Final Answer: $2\sqrt{3}/7$.**

1107
1108
1109
1110
1111
1112
1113
1114
1115
1116

Question: As shown in the diagram, the given solid consists of half of a cone and a triangular pyramid. The base of the cone is a circle with center O and radius 1, and the height of the cone satisfies $PO = 2$. The triangular pyramid P-ABC has its base ABC as an isosceles right triangle, where AB is the hypotenuse and coincides with the diameter of the cone's base circle. The base ABC lies on the same plane as the cone's base. Determine the angle between line PC and plane ABC.



Ground Truth: $\arctan 2$

Analysis: The generated auxiliary-line description **leaves the geometric configuration unchanged**. Consequently, the subsequent reasoning proceeds from a spurious auxiliary premise rather than from added geometric constraints, leading to an incorrect final answer.

1117
1118
1119
1120
1121
1122
1123

Solution: **[AUX]1. Construct auxiliary lines: Take the midpoint M of AB, and connect OM. [/AUX]** $\text{Determine perpendicularity: Since O is the center of the circle, } OM \perp AB. \text{ Also, } PO \perp OM \text{ because } PO \text{ is the height of the cone and } O \text{ is on the base circle. Since } AB \cap PO = O, \text{ and } AB \text{ and } PO \text{ are in plane PAB, we conclude } OM \perp \text{plane PAB.}$ $\text{Determine angle: The angle between line PC and plane ABC is the angle between line PC and its projection onto plane ABC. Since } OM \perp \text{plane PAB, the projection of point P onto plane ABC lies on line OM. The projection of line PC onto plane ABC is segment AC.}$ $\text{Calculate angle: To find the angle between line PC and plane ABC, we need to find the sine of the angle. Given } PO = 2 \text{ and } OA = 1, \text{ we have } AC = \sqrt{(PA^2 - AO^2)} = \sqrt{(2^2 - 1^2)} = \sqrt{3}.$ **Final Answer: $\arcsin(\sqrt{3} / 3)$**

Figure 8: Failure cases generated by GeoVLMath-7B.

1124
1125
1126
1127
1128
1129
1130
1131
1132
1133