GeoUni : A Unified Model for Generating Geometry Diagrams, Problems and Problem Solutions

Jo-Ku Cheng*, Zeren Zhang*, Ran Chen, Jingyang Deng, Ziran Qin, Jinwen Ma

* Equal contribution

程若谷 应用数学 马尽文

ABSTRACT

We propose GeoUni, the first unified geometry expert model capable of generating problem solutions and diagrams within a single framework in a way that enables the creation of unique and individualized geometry problems. Traditionally, solving geometry problems and generating diagrams have been treated as separate tasks in machine learning, with no models successfully integrating both to support problem creation. However, we believe that mastery in geometry requires frictionless integration of all of these skills, from solving problems to visualizing geometric relationships, and finally, crafting tailored problems. Our extensive experiments demonstrate that GeoUni, with only 1.5B parameters, achieves performance comparable to larger models such as DeepSeek-R1 with 671B parameters in geometric reasoning tasks. GeoUni also excels in generating precise geometric diagrams, surpassing both text-to-image models and unified models, including the GPT-4o image generation. Most importantly, GeoUni is the only model capable of successfully generating textual problems with matching diagrams based on specific knowledge points, thus offering a wider range of capabilities that extend beyond current models.

CONTRIBUTION

- We propose the first unified multi-modal geometry expert model, GeoUni, capable of solving geometry problems, generating precise geometric diagrams using both formal and natural language, and creating geometry problems based on knowledge points, supporting both English and Chinese.
- We propose Geo-MAGVIT, a module specifically designed for the tokenization of geometric diagrams. By introducing topo-structural awareness loss and text region loss, it significantly improves the accuracy of geometry structure and text reconstruction.
- We innovatively combine GRPO and LoRA to train the Geo-Reasoning-Adapter, which effectively boosts geometric reasoning capability and seamlessly integrates into the unified model architecture.
- We establish novel diagram generation evaluation metrics, which include the Geometric Semantic Matching Scores (GSMSs) and Geometric Pixel Matching Score (GPMS) to comprehensively evaluate the geometry generation task.

Overview of GeoUni

Scan the QR code for the full preprint.
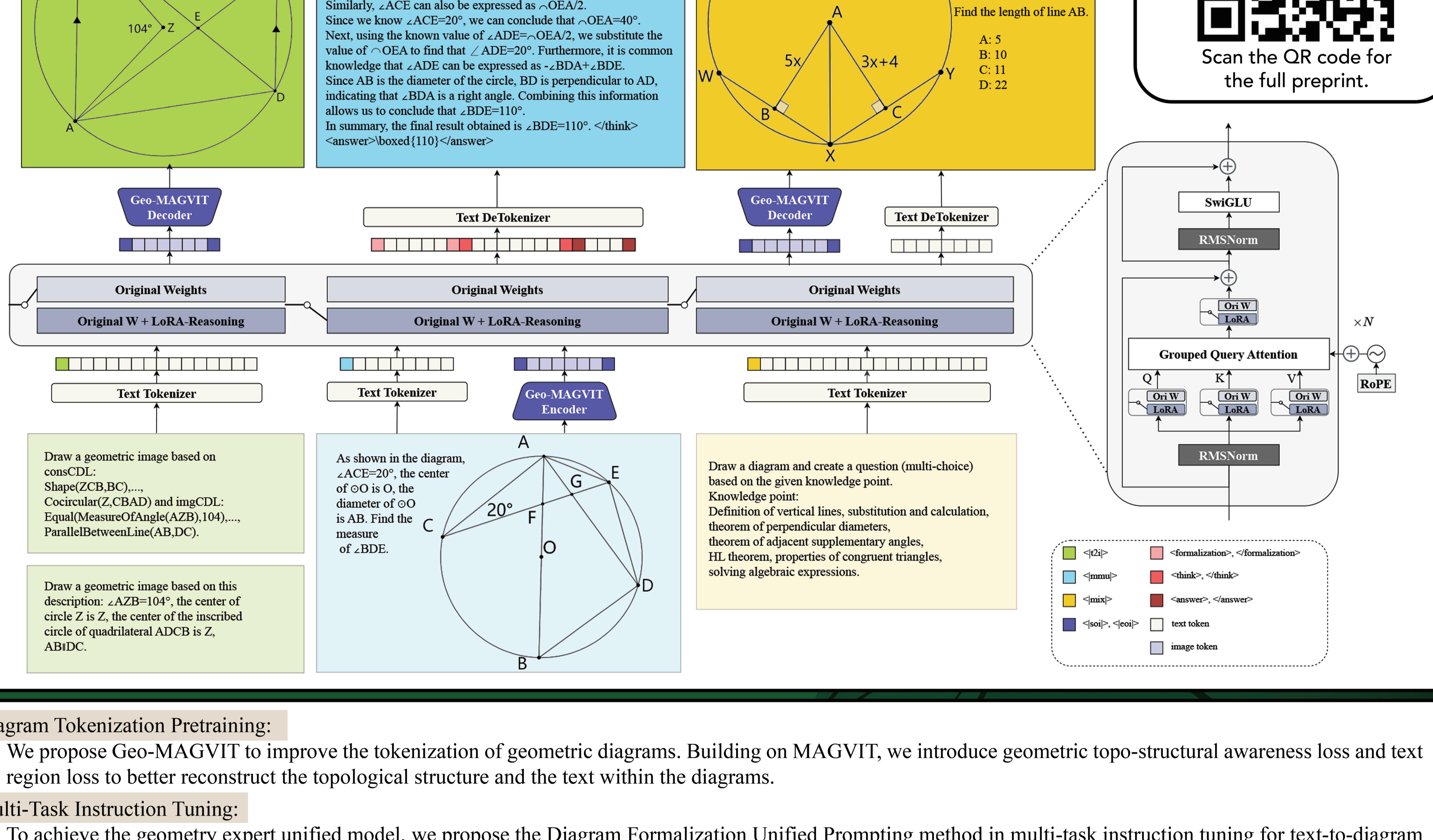
Diagram Tokenization Pretraining:
- We propose Geo-MAGVIT to improve the tokenization of geometric diagrams. Building on MAGVIT, we introduce geometric topo-structural awareness loss and text region loss to better reconstruct the topological structure and the text within the diagrams.
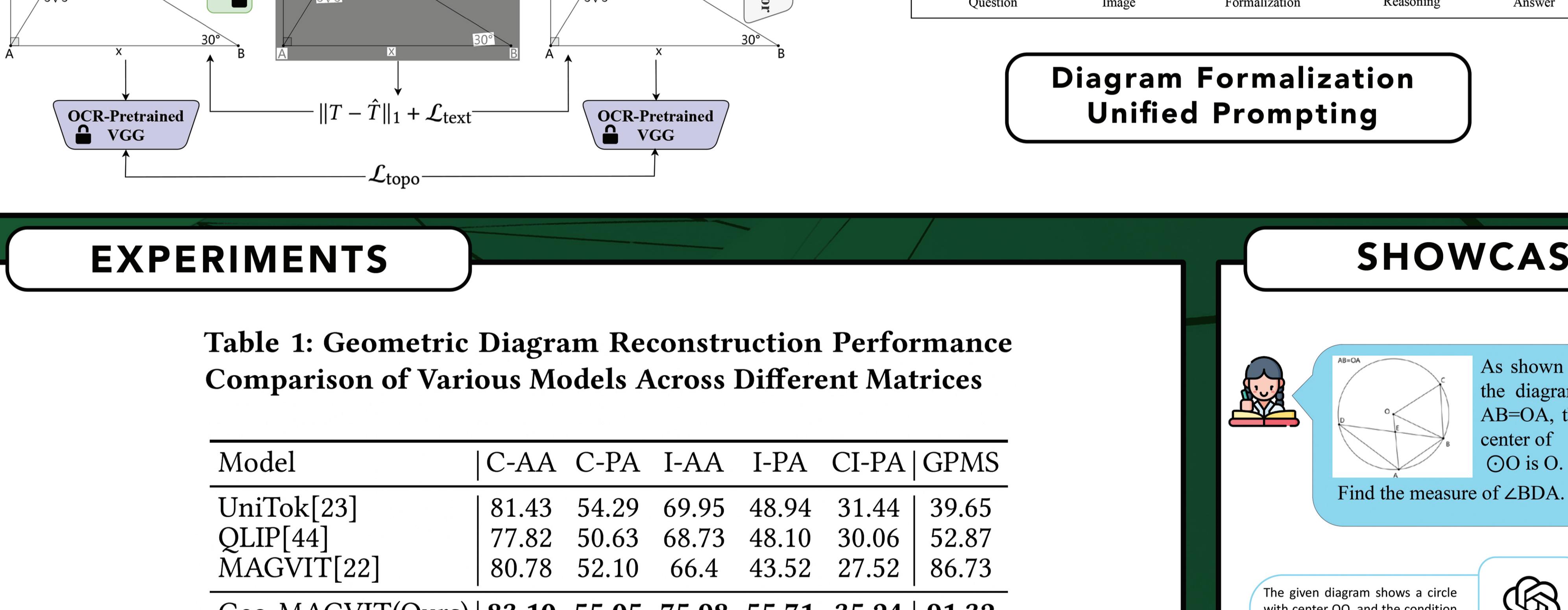
Multi-Task Instruction Tuning:
- To achieve the geometry expert unified model, we propose the Diagram Formalization Unified Prompting method in multi-task instruction tuning for text-to-diagram generation, problem solving, and problem generation, achieving next-token prediction training. This training phase equips GeoUni with the capability to accurately generate geometric diagrams, solve basic geometry problems, and generate problems based on knowledge points.
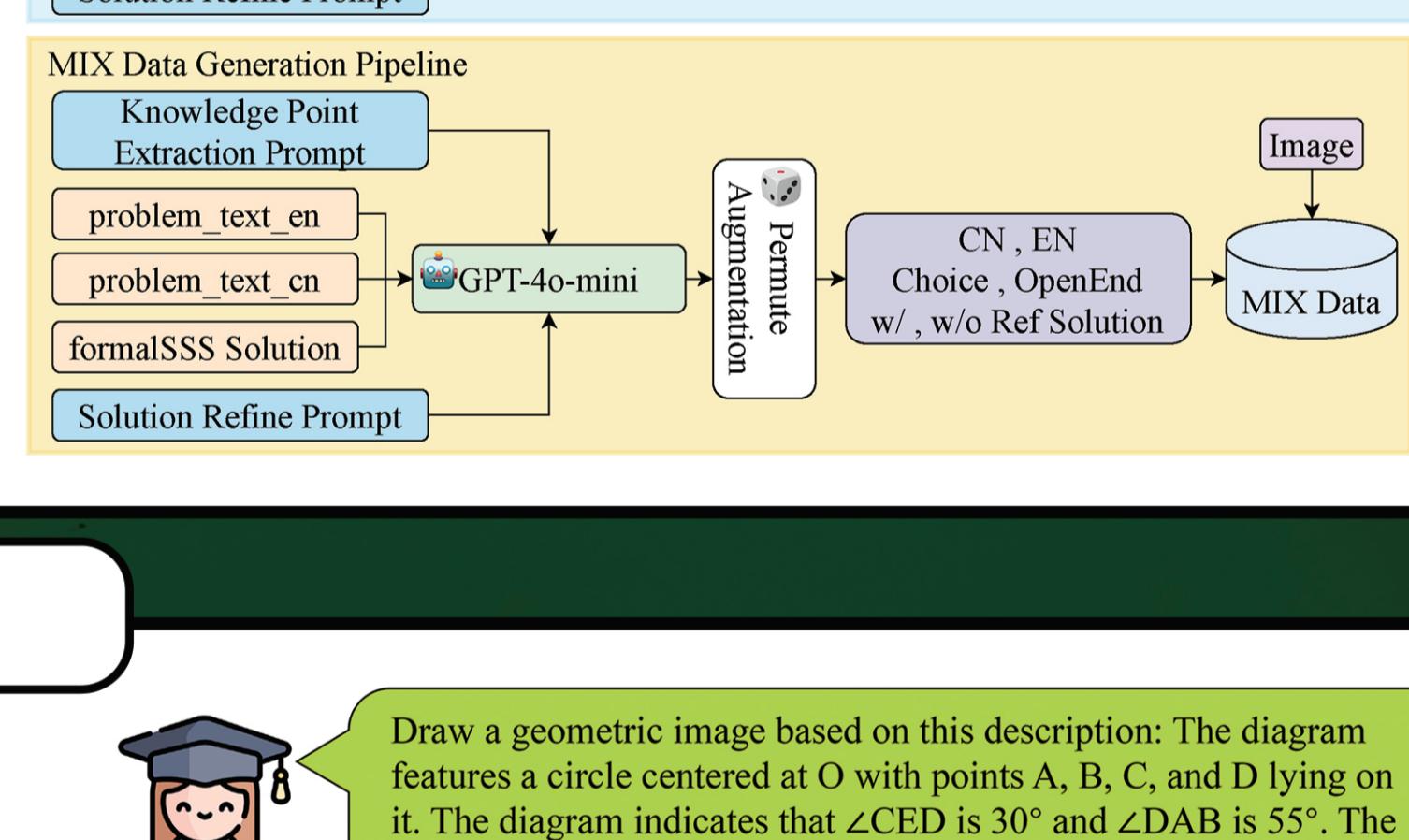
Reasoning Enhancement:
- We combine LoRA and GRPO to train the Geo-Reasoning-Adapter, which significantly improves the model's geometric reasoning ability while preserving its precise geometric diagram generation capability.

Overview of Geo-MAGVIT

Diagram Formalization Unified Prompting

Data Augmentation Pipeline

EXPERIMENTS

Table 1: Geometric Diagram Reconstruction Performance Comparison of Various Models Across Different Matrices

Model	C-AA	C-PA	I-AA	I-PA	CI-PA	GPMS
UniTok[23]	81.43	54.29	69.95	48.94	31.44	39.65
QLIP[44]	77.82	50.63	68.73	48.10	30.06	52.87
MAGVIT[22]	80.78	52.10	66.4	43.52	27.52	86.73
Geo-MAGVIT(Ours)	**83.10**	**55.05**	**75.98**	**55.71**	**35.24**	**91.32**

Table 2: Text-To-Diagram Performance Comparison of Various Models Across Different Matrices

Model	Construction/EN			Image/EN			Construction/CN			Image/CN		
	Caption	CDL	GPT	Caption	CDL	GPT	Caption	CDL	GPT	Caption	CDL	GPT
Show-o[38]	29.39	25.01	24.84	34.51	24.58	34.01	32.67	29.82	35.63	33.80	31.33	34.00
Janus-Pro-7B[7]	14.55	17.72	20.85	30.22	29.7	33.02	34.85	35.47	34.69	29.09	20.56	26.55
Anole-7B[8]	16.84	18.68	13.63	24.11	30.73	23.73	12.42	19.55	14.25	34.01	28.59	34.40
Emu3[36]	22.06	18.76	16.00	20.89	19.88	19.90	21.20	25.25	18.78	20.96	22.65	19.75
Unified-IO[19]	32.27	29.37	31.63	17.84	16.41	16.74	33.62	33.81	32.51	27.95	19.43	29.44
SEED-X[14]	34.20	31.78	27.84	14.80	10.47	12.30	30.78	31.40	27.40	15.60	11.37	14.89
PixArt-Σ[5]	21.69	17.94	17.14	19.35	19.44	15.39	28.55	23.67	28.29	17.89	21.40	16.45
SD-v1.5[29]	21.69	19.51	23.06	22.46	20.71	22.26	25.03	25.14	27.30	16.11	13.34	16.72
SDXL-Turbo[11]	12.94	11.12	14.76	13.86	15.93	11.76	30.36	17.86	29.62	19.52	10.84	18.38
DALL-E-2[27]	24.38	19.75	18.30	28.29	24.74	19.27	22.50	19.56	20.78	20.31	21.36	17.47
GeoUni-1.5B(Ours)	**73.00**	**73.43**	**72.41**	**78.46**	**79.65**	**77.53**	**73.72**	**73.00**	**72.77**	**79.40**	**79.54**	**77.82**
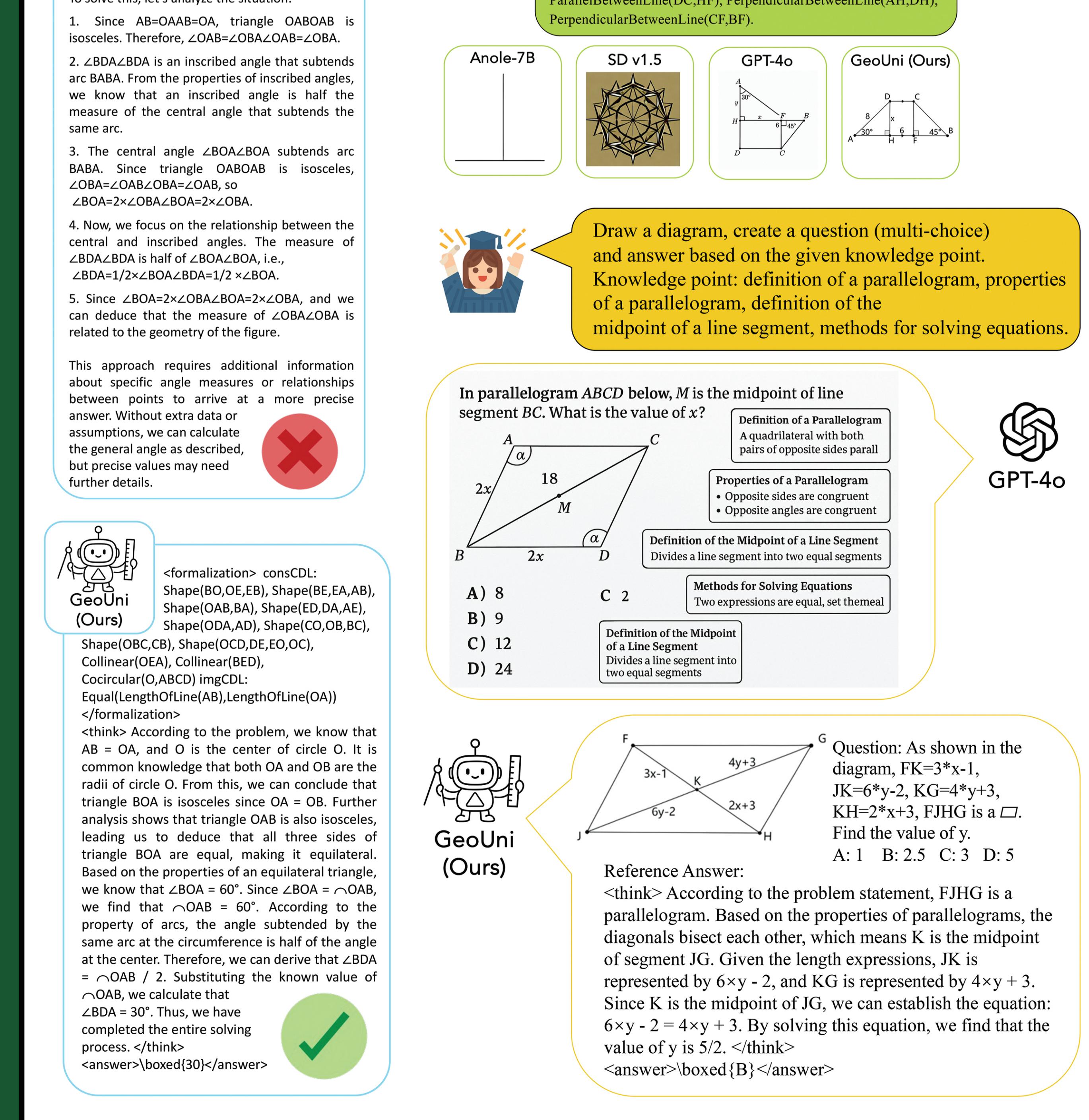
Table 3: Geometric Reasoning Performance Comparison of Various Models Across Different Matrices

Model	Formalgeo7K				Geometry3k				GeoQA			
	EN-C	EN-OE	CN-C	CN-OE	EN-C	EN-OE	CN-C	CN-OE	EN-C	EN-OE	CN-C	CN-OE
Show-o[38]	17.90	2.57	16.57	0.29	16.44	2.31	17.59	0.46	18.93	2.75	15.86	0.16
Chameleon-7B[32]	10.38	5.90	11.81	3.62	9.03	3.94	11.34	3.01	11.33	7.28	12.14	3.01
Janus-Pro-7B[7]	25.62	13.05	8.38	10.57	23.38	12.04	6.25	10.65	27.18	13.75	9.87	10.52
Emu3[36]	45.24	7.14	21.33	3.52	40.05	5.94	21.53	2.08	48.87	9.39	21.20	4.53
SEED-X[14]	13.90	1.33	7.24	3.43	11.81	0.69	5.09	1.85	15.37	1.78	8.74	4.53
GPT-4o[25]	47.52	20.10	49.71	28.48	56.09	23.38	52.08	28.70	45.31	17.80	48.06	28.32
G-LLaVA-13B[12]	47.81	14.10	-	-	5.56	34.95	-	-	56.80	20.06	-	-
Qwen2.5-VL-7B[2]	61.71	26.38	60.48	27.14	64.12	23.84	59.26	22.45	60.03	28.16	61.33	30.42
QWQ-72B[33]	51.62	16.57	20.19	7.14	50.69	23.61	18.52	7.41	52.27	22.15	21.36	6.96
Phi-4-Multimodal[24]	48.19	14.48	43.24	10.00	52.78	10.88	39.81	9.03	44.98	16.99	45.63	10.68
Qwen2.5-Math-1.5B[39]	48.95	24.12	47.43	20.86	45.60	20.14	47.45	21.99	51.29	26.90	47.41	20.06
DS-R1-Distill-Qwen-1.5B[9]	40.38	16.57	38.48	11.71	37.50	19.68	38.19	19.40	42.39	14.40	38.67	8.58
Qwen2.5-Math-72B[39]	70.67	37.90	67.90	37.62	69.44	36.81	65.05	33.33	71.52	38.67	67.90	40.61
DeepSeek-R1[9]	64.86	31.05	77.24	31.71	66.90	36.34	**75.23**	36.57	63.43	27.35	**78.64**	28.32
DeepSeek-V3[10]	70.29	33.46	65.71	26.38	71.30	35.66	63.19	28.01	69.58	30.32	67.48	25.24
GeoUni-1.5B(Ours)	**75.43**	**59.81**	73.52	**55.33**	**71.76**	**50.00**	69.68	**45.27**	**77.99**	**66.67**	76.21	**62.30**

Table 5: Impact of GRPO and formalization on our model's performance on the Formalgeo7K

Model	EN-C	EN-OE	CN-C	CN-OE
GeoUni (w/o GRPO)	53.81	39.90	55.52	39.33
GeoUni (w/o Formalization)	70.57	53.33	70.19	52.86
GeoUni	75.43	59.81	73.52	55.33

Table 4: Impact of Different Training Loss on Geo-MAGVIT

Topo	Text	C-AA	C-PA	I-AA	I-PA	CI-PA	GPMS
-	-	5.86	1.90	3.28	2.67	0.00	58.66
✓	-	43.24	26.38	18.78	8.00	4.48	71.11
✓	✓	**83.10**	**55.05**	**75.98**	**55.71**	**35.24**	**91.32**

SHOWCASE