MegaCQA: Construction and Benchmarking of a Large-Scale Visual Chart Question Answering Dataset

Xueqian Zheng, Tong Li, Zhou Liu, Qi Jiang, Xuhui Ma, and Guodao Sun

Zhejiang University of Technology

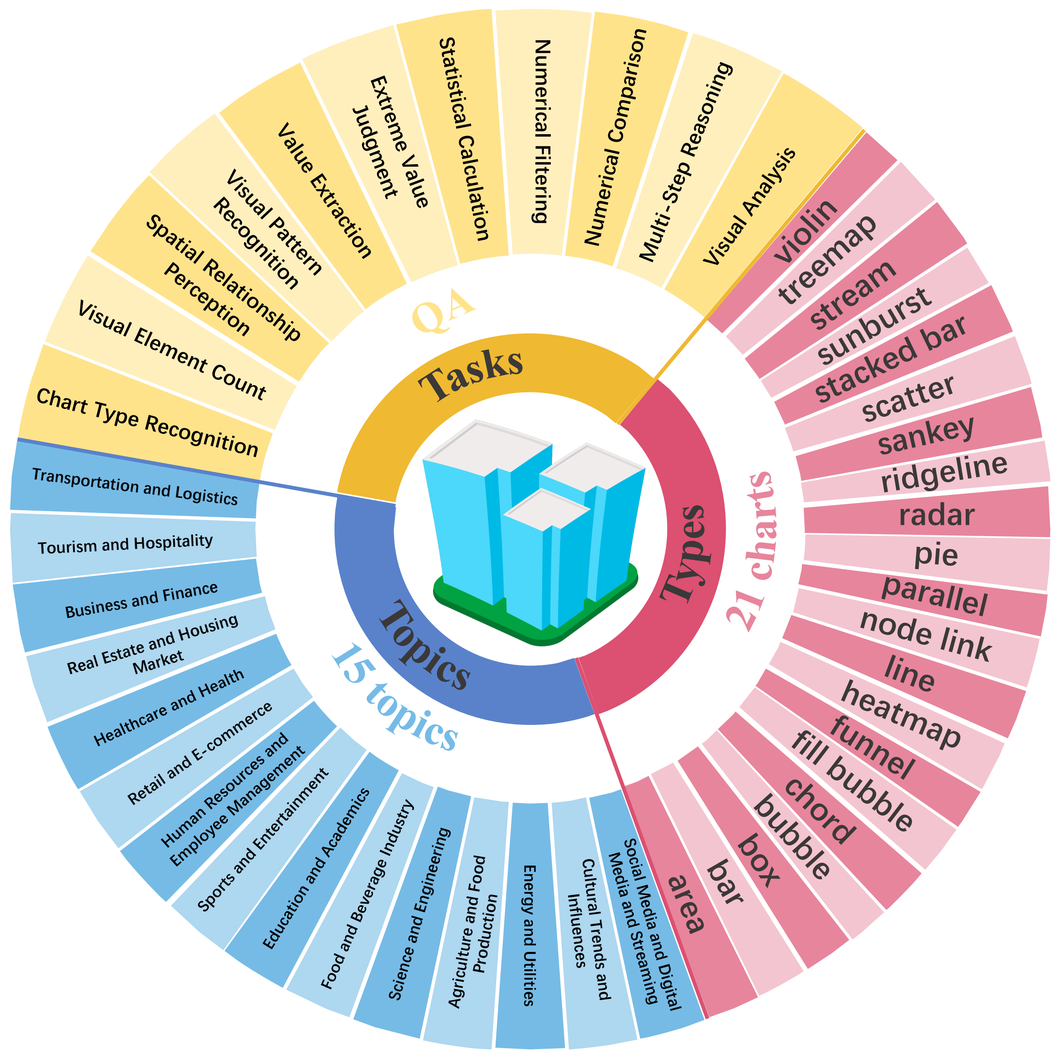
Introduction

As vision language models improve, multimodal understanding is increasingly applied to specialized fields, especially in chart question answering.

However, existing datasets are **small in scale**, **high data overlap**, **few chart types**, **single question formats**, and **restricted modalities**, making it difficult to fully evaluate a model's general reasoning abilities.

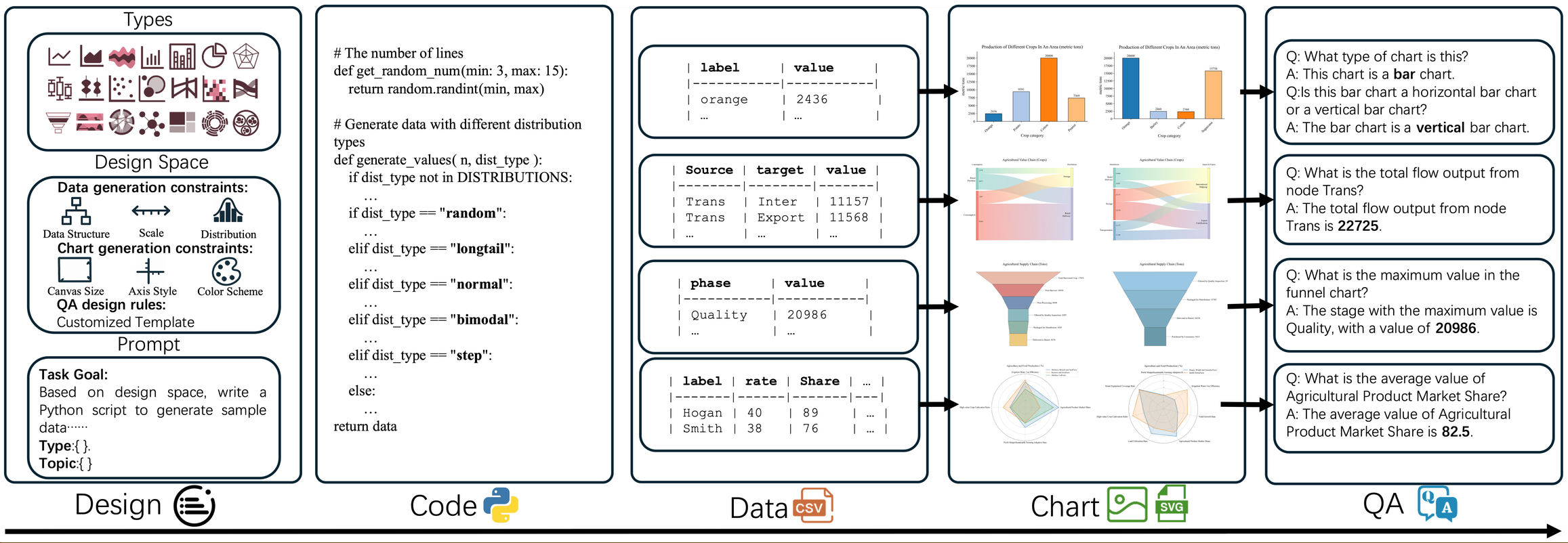
To address these limitations, we introduce a large-scale chart question answering dataset—**MegaCQA**. In summary, the contributions of this work are summarized as follows:

- We offer a wide range of chart types and question formats, enabling robust evaluation of model generalization to new scenarios.
- We use a multimodal data setup that not only supports visual reasoning tests but also makes it easy to analyze model performance in structured data understanding, numerical computation, and multi-step logical reasoning.
- We implement a multi-stage quality control process—combining hard constraints with soft constraints and both automated and manual checks—to ensure data accuracy and sample readability, providing a high-quality, reproducible benchmark for fair, systematic evaluation of complex chart understanding and reasoning models.



Pipeline

Dataset Construction Process



Comparison

Compared to existing datasets, MegaCQA represents a qualitative leap across several core metrics:

- It covers **21 chart types**, far exceeding its peers; includes a vast corpus of **260 K charts** and **5.74 M QA pairs.**
- Supports **11 question categories** with both **fixed** and **open** response formats, offering high task diversity.
- All samples are provided in **CSV, PNG, and SVG** modalities and undergo a multi-stage quality-control pipeline combining automated checks with expert review to ensure accuracy and readability.

Consequently, MegaCQA delivers a more comprehensive and reliable benchmark for evaluating models on structure parsing, numerical computation, and multi-step logical reasoning.

Dataset	Chart		QA Tasks		Answer Type		
	Types	Scale	Types	Scale	Closed	Fixed	Open
DVQA	1	300.0k	3	3487.4.k	√		√
FigureQA	3	140.0k	7	1800.0k	√		
PlotQA	3	224.3k	3	28900.0k	√	√	√
ChartQA	3	21.9k	4	32.7k		√	√
Chart-HQA	3	0.95k	4	2.2k			√
ChartInstruct	4	70.9k	-	42.4k	-	-	-
LEAF-QA	6	240.0k	2	2000.0k		√	√
ChartBench	8	66.6k	5	600.0k	√	√	
ChartLlama	8	11.0k	-	107.0k	-	-	-
ChartX	12	6.0k	-	48.0k		√	
NovaChart	14	47.0k	5	342.4k		√	√
MegaCQA(ours)	**21**	**260.0k**	**11**	**5740.0k**		√	√

ChinaVis 2025

浙江工業大学 ZHEJIANG UNIVERSITY OF TECHNOLOGY

zhengxueqian@zjut.edu.cn
litong@zjut.edu.cn
guodao@zjut.edu.cn