MIT: A intuitive tutor for mathematics reasoning

Yuchen Zhou Jialiang Xie Lifeng Hua Yinhui Liu Xiwen Huang Linfeng Fan Caoyang Zhang

Computer vision – Microsoft Project

Background

With the development of technology, generative video modeling has been applied in more and more fields.

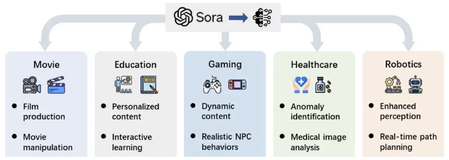

Figure 1. Applications of Sora

However, the application of 3d generative video screen modeling in mathematics is still relatively small. In the field of mathematics, 3d generative video screen modeling can intuitively and effectively help people understand mathematical problems and get the answers to mathematical problems. So we introduce a mathematics intuitive tutor (MIT). It combines a large language model with a generative video model to assist people in solving mathematical problems in both concrete and abstract perspectives.

Introduction

In the context of specific mathematical formulas and physical chemistry scenarios, such as neural network architectures or certain algorithms, there is a need for visualizing 3D videos to facilitate researchers' understanding. This application resembles the concept of AI for science (AI4Sci). The specific approach involves integrating existing models that excel in understanding mathematical reasoning and 3D video generation techniques, to effectively convey concepts to researchers.

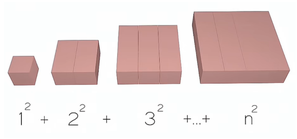

Figure 2. Visualization of mathematical problems

Research methodology

Our team envisions dividing this process into two main parts:

- **frontend technology implementation:** The frontend technology implementation primarily relies on existing mature large language models tailored to descriptions of mathematical formulas or physical chemistry environments.
- **backend technology implementation:** Meanwhile, the backend focuses on the visualization transformation of the output content from the large language model. In this part, considering that our resources are limited, we will mostly use the current 3D models like Latte[4] to create videos. Also, since the 3D models are trained by mostly real-life object datasets, so they naturally perform better at concrete examples while poor at abstract mathematics theories. So we also extend an interface on our backend specifically focusing on displaying abstract formulas.

Frontend technology implementation

In this section, our main task lies in choosing the appropriate large language model. Among many models, MathOctopus[2] is the most suitable one.

1. **Higher accuracy:** Notably, MathOctopus-13B reaches 47.6/100 accuracy which exceeds ChatGPT 46.3/100 on MGSM test set.

	Open-Source LLMs (7B Model)										
LLaMA 2-LoRA	27.6	4.0	12.0	2.0	10.4	18.4	16.8	7.6	11.2	3.2	11.3
LLaMA 2	43.2	5.2	22.4	3.2	37.2	32.4	34.4	15.2	28.0	4.8	22.6
RFT	44.8	2.8	16.8	2.4	33.6	34.0	34.0	6.8	29.2	2.0	20.6
MAmmoTH	49.6	2.4	17.2	3.6	33.2	32.4	32.8	10.8	26.0	4.8	21.3
WizardMath	47.6	3.4	22.4	2.0	30.4	34.8	30.4	24.0	30.8	4.0	23.0
MathOctopus^C	52.0	23.6	31.6	18.8	38.0	39.2	36.4	27.2	33.6	21.6	32.2
xRFT-MathOctopus^C	51.2	24.0	33.2	18.8	36.0	41.2	37.6	29.6	36.4	25.2	33.3
MathOctopus^P-LoRA	30.4	15.2	23.6	10.4	22.8	24.8	26.4	18.0	22.0	14.8	20.8
MathOctopus^P	52.4	**39.2**	38.4	28.8	44.8	42.4	43.6	36.0	39.6	34.4	40.0
xRFT-MathOctopus^P	**54.8**	38.4	**45.2**	**33.2**	**43.6**	**45.2**	**38.0**	**35.6**	**48.4**	**36.4**	**41.9**

Figure 3. Accuracy Comparison

2. **Recognize different languages:**It can handle ten languages such as English, Chinese, Japanese, Russian, etc.

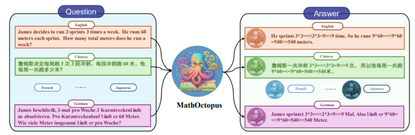

Figure 4. Processing of different languages

Backend technology implementation

The first part of it, We chose Latte[4] as our back end model. Latte is a simple general, and efficient video diffusion method and is widely used by T-to-V open-source frameworks. In general, Latte consists of two main modules: a pre-trained VAE and video DiT.

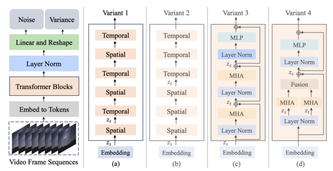

Figure 5. The pipeline of Latte for video generation

It composes of three parts: video compression, spatiotemporal modeling and features restore.

VAE : Latent Diffusion Models LDMs learn the data distribution through two key processes: diffusion and denoising. Notably, the denoising process uses model ϵ_θ and the following equation:

$$\mathcal{L}_{simple} = \mathbb{E}_{\mathbf{z}\sim p(z),\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(\mathbf{z}_t, t)\|_2^2\right]. \quad (1)$$

Apart from the text2video model that we used to generate concrete examples, we also used OpenAI API to build abstract mathematics equations. We use a OpenAI API to automatically generate python codes to reach our goals to display the effects.

Project results and summary

In our project, we have created a comprehensive mathematics tool for intuitive tutoring. Both in the concrete and abstract aspects, we have provide users with a simple illustration of the problem, helping them to achieve a better understanding. **Our code and research results are now available at https://github.com/ErwinZhou/Mathematics-Intuitive-Tutor.**

Concrete Examples

We have used MathOctopus[2] to gain a precise understanding of the math problems. And then we feed it to the input of Latte[4] to generate real-life examples for simple mathematics tutoring. It efficiently works for basic scenarios.

Abstract Mathematics Formulas

We have used OpenAI API to automatically generate python code to display a 3D illustration of abstract mathematics formulas and thesis. It can lead to a more eye-catching and realistic understanding for simple usage.

Future developments and challenges

Even though we have created a well-rounded application of the Image Synthesis for mathematics tutoring, it still have some shortcomings due to our short of time and computing resources. We hoping to overcome these in the near future:

- **Improving Visual Effects:** Although we have achieved a temporary effect to display simple scenes. It still lacks in performing smoothly and more attached to the content.
- **Integrating both ends:** Although we have created usage for both concrete and abstract aspects, we have used different front end for both problems. It will be great if we can use the MathOctopus[2] as the overall front end to replace OpenAI API to achieve a more coherent structure.

References

[1] Janice Ahn, Rishu Verma, Renze Lou, Di Liu, Rui Zhang, and Wenpeng Yin.
Large language models for mathematical reasoning: Progresses and challenges.
arXiv preprint arXiv:2402.00157, 2024.

[2] Nuo Chen, Zinan Zheng, Ning Wu, Linjun Shou, Ming Gong, Yangqiu Song, Dongmei Zhang, and Jia Li.
Breaking language barriers in multilingual mathematical reasoning: Insights and observations.
arXiv preprint arXiv:2310.20246, 2023.

[3] Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, et al.
Solving quantitative reasoning problems with language models.
Advances in Neural Information Processing Systems, 35:3843–3857, 2022.

[4] Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Ziwei Liu, Yuan-Fang Li, Cunjian Chen, and Yu Qiao.
Latte: Latent diffusion transformer for video generation.
arXiv preprint arXiv:2401.03048, 2024.

[5] Claude E. Shannon.
A mathematical theory of communication.
Bell System Technical Journal, 27(3):379–423, 1948.

[6] Ryutaro Yamauchi, Sho Sonoda, Akiyoshi Sannai, and Wataru Kumagai.
Lpml: llm-prompting markup language for mathematical reasoning.
arXiv preprint arXiv:2309.13078, 2023.