Development of Text-Guided Control for Temperature Distribution and Flow Patterns in SiC Solution Growth using Language Models

Kota Takahara[1], Kentaro Kutsukake[1,2], Shunta Harada[1,2] and Toru Ujihara[1,2]
[1]Guraduate School of Engineering, Nagoya University, Japan
[2] Institute of Materials and Systems for Sustainability (IMaSS), Nagoya University, Japan
takahara.kota.f1@s.mail.nagoya-u.ac.jp

Introduction

Recently, large language models (LLMs) have achieved remarkable progress and are increasingly applied to molecular design and inorganic crystal structure design[1], attracting attention as powerful tools to accelerate materials development. In the SiC solution growth process targeted in this study, the design of temperature and flow distributions inside the crucible is essential for the growth of high-quality crystals. This study aims to enable such design through intuitive linguistic instructions from domain experts, rather than relying solely on increasingly complex machine learning and optimization techniques. To this end, we aimed to construct a paired dataset for contrastive learning of text–image relations. Furthermore, we analyzed the patterns of hallucinations that multimodal LLMs exhibit when interpreting fluid distributions.

Experimental Procedures

Firstly, CFD simulations were used to obtain 600 temperature and flow distributions inside the crucible. Based on these results, we constructed a high-quality text dataset without hallucinations by generating descriptive texts using GPT-4o. To suppress hallucinations during text generation, we examined the introduction of ReAct framework, which integrates reasoning and action, allowing the model to produce explanations step by step in response to image inputs. The prompts included background information on the relative positions of the crucible center, sidewall, and seed crystal. For evaluation, we specified three aspects and assessed whether the generated descriptions correctly represented these elements, thereby validating the quality of the dataset.

Results and Discussion

Figure 1 shows the accuracy of directional and thermal descriptions in LLM outputs. When image inputs were used, hallucinations—particularly misidentifications between upward and downward flows—were frequently observed in directional explanations. This can be attributed to the fact that current multimodal LLMs are typically pretrained with CLIP[2], which relies on large-scale image–text pairs but contains limited information on fluid "direction." As a result, the models have not been trained to capture such fine-grained distinctions. In contrast, with the ReAct, directional and thermal judgments are supported with external computation tools, which greatly reduces the occurrence of hallucinations

Furthermore, Figure 2 presents the log probability when outputting directional tokens. In particular, when generating the token 6 meaning "upwards" the log probability remained low and unstable, leading to inconsistent outputs.

These findings indicate that CLIP-based training is insufficient for generating accurate textual explanations of fine differences in temperature and flow distributions. Therefore, improvements in dataset design for contrastive learning and modifications in patch segmentation within Vision Transformers are required to overcome these limitations.

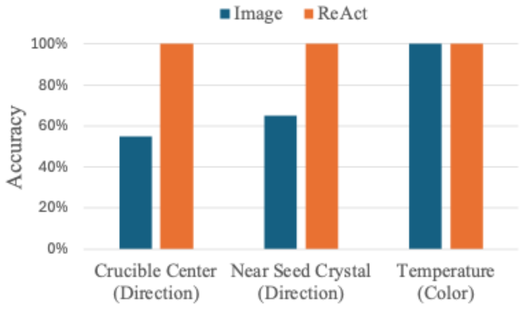

Figure 1 Accuracy of directional and thermal descriptions

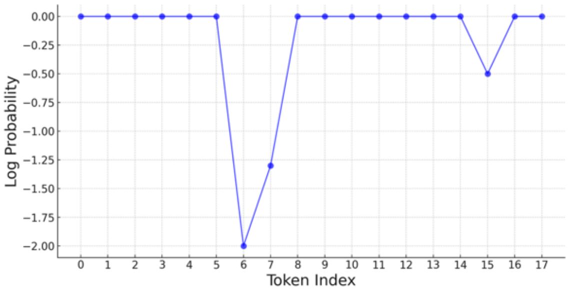

Figure 2 Log probability of each tokens

References

[1] K. Ozawa, T. Suzuki, S. Tonogai and T. Itakura, Sci. Technol. Adv. Mater.: Methods 4, 2406219 (2024).

[2] A. Radford, J. W. Kim, C. Hallacy et al., Proc. Mach. Learn. Res. 139, 8748-8763 (2021).