

Peer Review

Review of: "Benchmark Evaluations, Applications, and Challenges of Large Vision Language Models: A Survey"

Jiaxin Qi¹

1. Nanyang Technological University, Singapore

This survey provides an excellent summary of the majority of models, tasks, and datasets in the visual-language (VL) domain. The content is meticulously organized, and corresponding open-source code is provided, which has gradually gained attention. Overall, the quality of this paper is very high, covering a broad range of topics in a clear and understandable manner. It is well-suited as a survey to help readers understand the latest developments in the field of visual language.

Here are some suggestions: 1. Although the paper describes the general encoding process of VL models, this description remains too simplistic given the complexity and diversity of VL tasks. In particular, the integration of vision and language can vary slightly across different tasks. The authors should consider highlighting these differences to better aid readers in understanding each type of VL task.

2. With the proliferation of models, benchmarks, and datasets, many research teams claim that their models are the best. For new researchers entering the field, how to compare the strengths and weaknesses of different models is a significant challenge. The authors are encouraged to delve deeper into this issue, providing clearer and more objective evaluation methods for current models, thereby helping readers form a more accurate understanding.

3. The six challenges identified by the authors for VLMs are highly constructive. However, in the VLM domain, it is crucial to discuss how to integrate existing evaluation standards, especially considering that some datasets are outdated or of low quality. Additionally, selecting universally recognized benchmarks is an important issue that needs further exploration. The authors should address these concerns to promote standardization and consistency within the field.

Declarations

Potential competing interests: No potential competing interests to declare.