先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺　太郎 (教授))		
学籍番号	2311224	提出日	令和 7年 1月 21日
学生氏名	林　和樹		
論文題目	Vision Large Language Model Evaluation		

要旨

Large-scale Vision-Language Models (LVLMs) demonstrate advanced capabilities in text generation and comprehension from images and instructions, enhancing tasks like Visual Question Answering (VQA), image captioning and image recognition.
However, practical applications face challenges in integrating complex image-based knowledge, providing clear explanations, and generating context-appropriate text from various perspectives.
This work introduces two novel evaluation frameworks to evaluate these aspects.
Firstly, the Artwork Explanation Generation task evaluate how well LVLMs understand and integrate the knowledge required to explain images and the complex relationships between various elements.
LVLMs are tested on generating explanations using both images and titles, and images alone, to evaluate their language and vision-based knowledge.
Results show that LVLMs struggle to integrate visual and textual information and derive sufficient knowledge from images alone.
Secondly, the Image Review Rank (IRR) framework evaluates how well LVLMs align with human interpretations by assessing their ability to identify the most contextually appropriate critiques for an image.
This task examines whether LVLMs can identify the most context-appropriate interpretations of images, acknowledging that image interpretation varies by context. Experiments show that while LVLMs perform consistently across languages, their correlation with human annotations is low, revealing deficiencies in understanding human reasoning and capturing context.