UNIVERSITY OF SCIENCE

ADVANCED PROGRAM IN COMPUTER SCIENCE

COMMENTS OF THESIS'S REVIEWER

(Research)

Thesis title:

SMART INTERACTIVE RETRIEVAL OF VISUAL DATA VIA SEMANTIC UNDERSTANDING

Students: **Hoàng Xuân Nhật** (18125042) – **Nguyễn E Rô** (18125046)

Advisor: **Assos. Prof. Trần Minh Triết**

1. Research Topics and Ideas:

The main objective of the thesis is to apply AI to the search system to make it smarter, explainable and maintaining scalability.

The idea is to use multimodal interaction with visual-language model with pretrained model to make the system to be more generic and could adapt with any kind of queries of the users.

2. Research Methods:

The author investigated on many state-of-the-art models for visual-language search such as CLIP and indexing such as FAISS and pretrained models for NLP and CV such as BERT, ResNet.

Based on these, the authors propose the system FIRST and VLFormer to improve the retrieval system in both accuracy, scalability and ability to explain in steads of showing the final results like a black box.

3. Contributions:

The thesis has three main contributions:

(i) The adoption of a vector database to store the embeddings generated by AI models for use in searching, besides using an ordinary database for storing

metadata. This approach prevents the AI model from becoming the bottleneck of the system, enabling querying speed comparable to non-AI systems.

(ii) The extensive experience from varying users to form a list of guidelines, or principles, for users to keep in mind while searching. This gives users some sense of direction while still being flexible enough to adapt to different situations.

(iii) A novel referring expression segmentation module to precisely point out the object or concept referred to in the image. This strengthens the model's credibility in all scenarios, especially in some critical ones such as medical uses.

4. Report:

The thesis report is represented in a well format with 5 chapters and 74 pages. The structure of the report is suitable with a scientific document.

5. Presentation:

The authors are very confident in representation at the defense session. The slide is prepared in good format, but still have some minor mistakes.

6. Publications and/or realworld applications:

- The authors published 4 papers (2 accepted, 2 in review) and 1 workshop paper (CVPR 2022)
- The system was also used successfully in an international interactive video browser showdown contest organized by Multimedia Modeling conference 2022, as well as in Lifelog Search Challenge at ACM International Conference on Multimedia Retrieval 2022.

Rank: *Outstanding* **10 + 1** (bonus)

Ho Chi Minh city, 26 August 2022

Reviewer

Nguyễn Vinh Tiệp