VIMVQA - Object-based Reasoning Vietnamese Medical Visual Question Answering

Lê Trọng Đại Trường - 22521576
Nguyễn Hữu Hoàng Long - 22520817

Tóm tắt

- Lớp: CS519.O21.KHTN
- Link Github: https://github.com/ShouyiLeee/CS519.O21.KHTN
- Link YouTube video: https://youtu.be/JFC7Y3VxBn8
- Ảnh + Họ và Tên của các thành viên:

Lê Trọng Đại Trường

Nguyễn Hữu Hoàng Long

Giới thiệu

- Med-VQA là sự kết hợp giữa Computer Vision và NLP, sử dụng đầu vào là một hình ảnh y khoa (Medical image) và câu hỏi liên quan, được kỳ vọng sẽ cung cấp câu trả lời chính xác.

Q: What does the ct scan of thorax show?
A: bilateral multiple pulmonary nodules

Q: Is the lesion associated with a mass effect?
A: no

- Ở Việt Nam, do điều kiện còn hạn chế nên lĩnh vực này chưa thực sự phát triển và còn rất nhiều tiềm năng để khai thác.

Giới thiệu

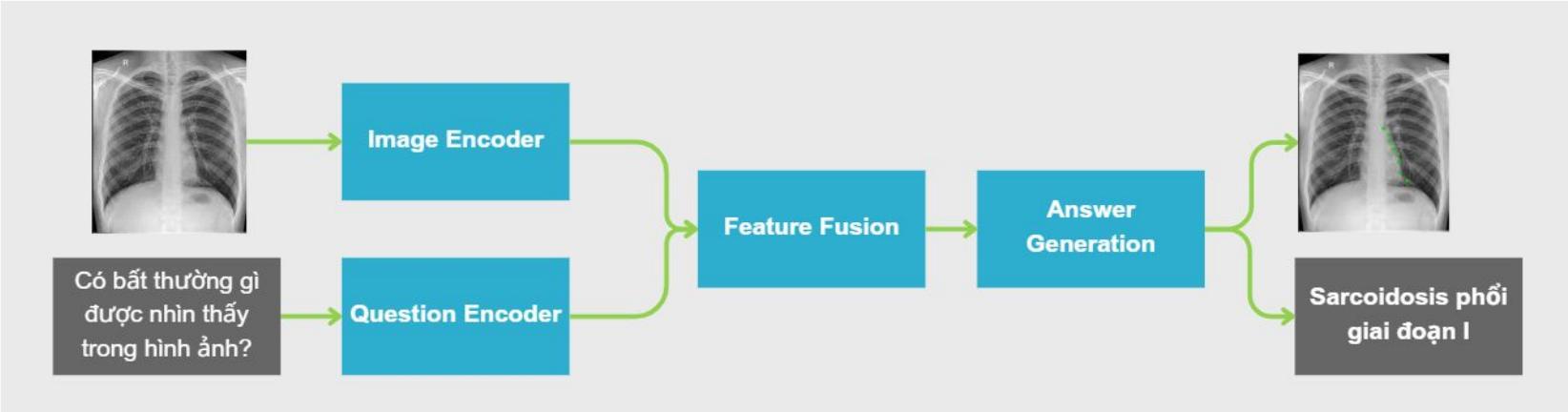
- Trong dự án nghiên cứu này, chúng tôi sẽ đề xuất một mô hình MedVQA dành cho ngôn ngữ tiếng Việt có tên là VIMVQA với giả thiết sẽ tăng cường độ chính xác của mô hình trên dữ liệu tiếng Việt.
- Song song với đó, chúng tôi sẽ xây dựng một kiến trúc mới cho phép mô hình đưa ra các bằng chứng luận lý (Reasoning evidence) để giải thích kết quả.
- Ngoài ra, chúng tôi còn xây dựng một bộ dữ liệu tiếng Việt về MedVQA có tên là VIMVQA-Data.

Mục tiêu

- Thu thập, xây dựng bộ dữ liệu tiếng Việt VIMVQA-Data nhằm phục vụ cho việc nghiên cứu, đánh giá.
- Nghiên cứu, đề xuất phương pháp mới cho bài toán MedVQA trong ngôn ngữ tiếng Việt. Trong đó tìm ra bộ Encoder-Decoder mới, có hiệu năng cao hơn trong ngôn ngữ tiếng Việt so với các bộ Encoder-Decoder đã có và thay đổi kiến trúc để có thể trực quan hóa bằng chứng cho quá trình đưa ra kết quả.
- Xây dựng ứng dụng minh họa có giao diện thân thiện, dễ sử dụng.

Nội dung và Phương pháp

- Tiến hành khảo sát các phương pháp đã có cho bài toán MedVQA. Chạy thực nghiệm, phân tích các điểm mạnh và yếu của các phương pháp để làm cơ sở cho các thử nghiệm cải tiến. Từ đó, đề xuất ra những kĩ thuật, cách tiếp cận mới để khắc phục các điểm yếu và đồng thời tăng hiệu năng khi áp dụng với ngôn ngữ tiếng Việt.



Nội dung và Phương pháp

- Quá trình cải tiến sẽ tập trung vào việc nâng cấp bộ Encoder-Decoder.Thêm vào đó, kết hợp việc sử dụng các kĩ thuật như GradCAM để làm bằng chứng để xác thực cho câu trả lời.
- Nghiên cứu và tìm hiểu về các bộ Encoder dành cho tiếng Việt có sẵn như PhoBERT, Vietnamese-SBERT, ViT5 hoặc xây dựng một bộ Encoder mới. Ngoài ra, tìm hiểu thêm các kĩ thuật để tăng hiệu quả cho việc Xử lý ngôn ngữ tự nhiên như Named Entity Recognition (NER).
- Thu thập, tiền xử lý và xây dựng một bộ dữ liệu MedVQA cho ngôn ngữ tiếng Việt có tên là VIMVQA-Data.

Nội dung và Phương pháp

- Tìm hiểu về các độ đo, phương pháp đánh giá trong MedVQA như Accuracy, BLEU, AUC-ROC.
- Thực nghiệm, đánh giá các phương pháp đã đề xuất trên bộ dữ liệu VIMVQA-Data và các bộ dữ liệu khác. Từ đó rút ra kết luận và đề xuất VIMVQA với kì vọng là mô hình có hiệu năng tốt hơn hẳn so với các mô hình trước trong bài toán MedVQA với ngôn ngữ tiếng Việt.
- Tìm hiểu điều kiện cũng như ứng dụng thực tế. Xây dựng một chương trình thực nghiệm thỏa mãn các ràng buộc và đáp ứng được nhu cầu của người sử dụng. Theo dõi, thu thập số liệu phục vụ cho nhu cầu nâng cấp, bảo trì sau này.

Kết quả dự kiến

- Hiểu rõ ý tưởng, cách hoạt động cũng như ưu, nhược điểm của các phương pháp đã có cho bài toán MedVQA.
- Hiểu rõ các độ đo cũng như phương pháp đánh giá được sử dụng trong bài toán MedVQA cho ngôn ngữ tiếng Việt.
- Bảng thống kê và so sánh chi tiết của toàn bộ phương pháp qua các độ đo đã đề cập.
- Mô hình VIMVQA được đề xuất cho kết quả tốt hơn so với phần còn lại và có thể cung cấp các bằng chứng luận lý cho kết quả đưa ra.
- Ứng dụng thực tế minh họa chi tiết.

Tài liệu tham khảo

- [1] Louisa Canepa, Sonit Singh, Arcot Sowmya: Visual Question Answering in the Medical Domain. DICTA 2023: 379-386
- [2] Zhihong Lin, Donghao Zhang, Qingyi Tao, Danli Shi, Gholamreza Haffari, Qi Wu, Mingguang He, Zongyuan Ge: Medical visual question answering: A survey. Artif. Intell. Medicine 143: 102611 (2023)
- [3] Xiaoman Zhang, Chaoyi Wu, Ziheng Zhao, Weixiong Lin, Ya Zhang, Yanfeng Wang, Weidi Xie: PMC-VQA: Visual Instruction Tuning for Medical Visual Question Answering. CoRR abs/2305.10415 (2023)
- [4] Junnan Li, Dongxu Li, Silvio Savarese, Steven C. H. Hoi: BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. ICML 2023: 19730-19742
- [5] Khiem Vinh Tran, Hao Phu Phan, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen: ViCLEVR: A Visual Reasoning Dataset and Hybrid Multimodal Fusion Model for Visual Question Answering in Vietnamese. CoRR abs/2310.18046 (2023)
- [6] Quan Van Nguyen, Dan Quang Tran, Huy Quang Pham, Thang Kien-Bao Nguyen, Nghia Hieu Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen: ViTextVQA: A Large-Scale Visual Question Answering Dataset for Evaluating Vietnamese Text Comprehension in Images. CoRR abs/2404.10652 (2024)