

Enhancing Vision-Language Models for Visual Question Answering with Low-Resources via Prompt-Based Learning.

Trần Vĩnh Khiêm^{1,2}

¹ Trường ĐH Công nghệ Thông tin, thành phố Hồ Chí Minh, Việt Nam

² Đại học Quốc gia Thành phố Hồ Chí Minh, Việt Nam

What ?

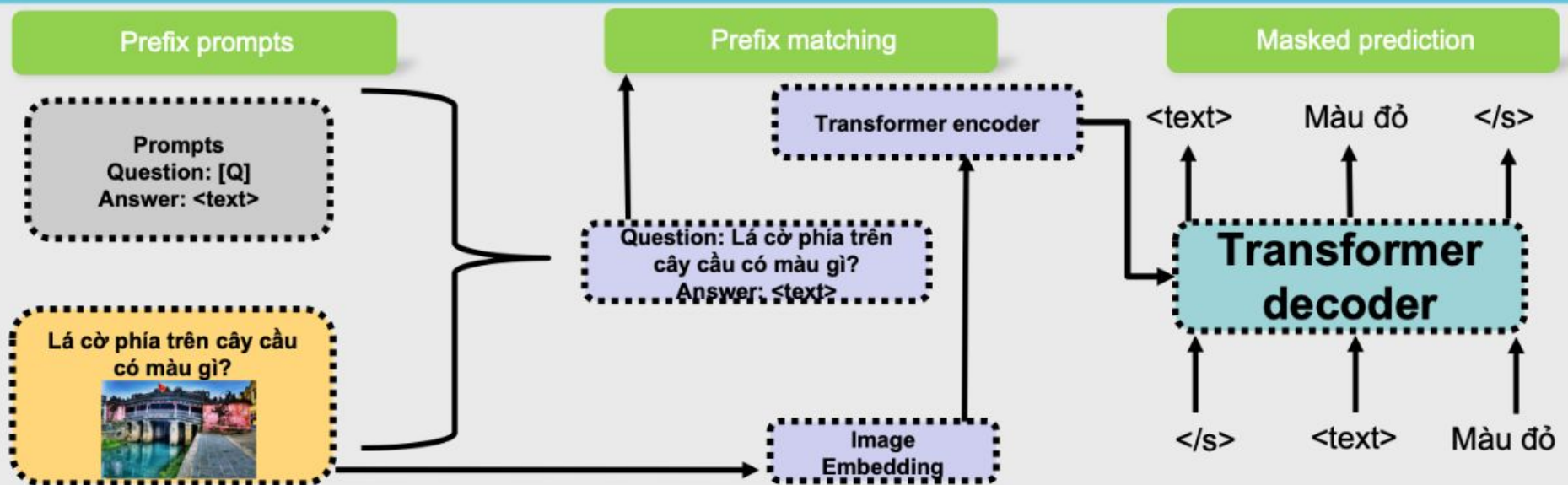
Đề xuất phương pháp nhằm tiết kiệm phần cứng tính toán và đảm bảo hiệu suất mô hình, mục tiêu bao gồm:

- Đề xuất phương pháp transformer sử dụng prefixLM và MaskedLM.
- Tiến hành đánh giá mô hình với nhiều prompts khác nhau.
- So sánh mô hình với các kích thước dữ liệu huấn luyện khác nhau.
- Thử nghiệm mô hình trên các tác vụ zero-shot và few-shot

Why ?

- Việc thu thập dữ liệu và gán nhãn bởi con người tốn kém và chậm
- Các mô hình ngôn ngữ - hình ảnh lớn gần đây quá lớn để triển khai trên máy tính vừa và nhỏ.
- Việc thiết kế prompt không hiệu quả làm giảm độ chính xác của mô hình
- Các prompt đã được chứng minh tầm quan trọng và hiệu quả trong các tác vụ sử dụng ít dữ liệu trên văn bản

Overview



Description

1. Đề xuất mô hình PromptVQA

- Khảo sát trên các hội nghị A/A* như CVPR, ACL, ICCV, EMNLP
- Đề xuất mô hình sử dụng PrefixLM và MaskedLM
- Tiến hành so sánh với các mô hình LLMs (mô hình ngôn ngữ lớn)



2. Thử nghiệm nhiều loại prompts và phân tích hiệu suất

- Nghiên cứu các loại prompts như hand-crafted prompts, noisy prompts, irrelevant prompts
- Nghiên cứu tác động của các loại prompt khác nhau trong các tác vụ zero-shot và few-shot



- Đề xuất prompt mẫu theo các dạng đã tìm hiểu được.
- Sử dụng prompt mẫu và tiến hành đánh giá hiệu quả so với câu hỏi thông thường trên mô hình đề xuất và các mô hình LLMs.

3. Huấn luyện mô hình với các kích thước dữ liệu khác nhau

- Tiến hành huấn luyện các mô hình với các kích thước dữ liệu khác nhau và các prompt khác nhau
- So sánh hiệu suất mô hình đề xuất so với các mô hình LLMs

4. Tiền huấn luyện và đánh giá trên các tác vụ khác nhau sử dụng zero-shot và few-shot

- Nghiên cứu các mục tiêu huấn luyện khác nhau ảnh hưởng đến các tác vụ khác nhau của VQA.
- Tiến hành đánh giá mô hình đề xuất với các mục tiêu khác nhau
- Đánh giá và so sánh mô hình đề xuất

Kết quả mong đợi

- Phương pháp đề xuất có kết quả cao hơn hoặc bằng với các mô hình LLMs ở thời điểm hiện tại trên độ đo Accuracy và cần tiêu tốn tài nguyên hơn.
- Các mẫu prompts có thể làm gia tăng hiệu suất của mô hình
- Mô hình đề xuất có hiệu suất cao kể cả khi huấn luyện với dữ liệu hạn chế
- Mô hình đề xuất có khả năng tổng quát hoá tốt trên các tác vụ VQA.