

TĂNG CƯỜNG HIỆU QUẢ CỦA MÔ HÌNH SINH MÔ TẢ CHO HÌNH ẢNH BẰNG CÁCH KẾT HỢP MÔ HÌNH NGÔN NGỮ LỚN VÀ MÔ HÌNH THỊ GIÁC

Trần Ngọc Thiện

Trường ĐH Công nghệ thông tin

What ?

- **Bài toán:** Image captioning - Tự động tạo mô tả văn bản cho hình ảnh.
- **Giải pháp:** Kết hợp mô hình ngôn ngữ lớn (LLM) và mô hình thị giác để cải thiện chất lượng mô tả.

Why ?

- **Hạn chế của các phương pháp hiện tại:** Các phương pháp truyền thống như encoder-decoder và transformer-based còn hạn chế về khả năng nắm bắt ngữ cảnh, tạo mô tả phong phú và xử lý thông tin hình ảnh.
- **Cần thiết:** Cần một giải pháp hiệu quả hơn để tạo ra các mô tả hình ảnh chính xác, chi tiết và phù hợp với ngữ cảnh.

Overview

Đề tài tập trung vào việc phát triển một mô hình mới kết hợp sức mạnh của LLM (khả năng sinh văn bản tự nhiên) và mô hình thị giác (khả năng phân tích hình ảnh) để giải quyết bài toán image captioning. Mô hình này được kỳ vọng sẽ tạo ra các mô tả hình ảnh chất lượng cao hơn, đáp ứng tốt hơn nhu cầu thực tế.

Description

Quy trình:

- Khảo sát các mô hình thị giác và mô hình ngôn ngữ lớn phù hợp.
- Xây dựng mô hình:
- Huấn luyện mô hình trên bộ dữ liệu KTVIC.
- Đánh giá mô hình trên các thang đo BLEU, METEOR, ROUGE.
- So sánh hiệu suất của mô hình với các phương pháp baseline.

Mục tiêu:

- Xây dựng một mô hình hiệu quả cho bài toán image captioning.
- Cải thiện chất lượng các mô tả hình ảnh.
- Mở rộng ứng dụng của công nghệ trong thực tế.