

Signboard text detection and recognition in streaming video

Nguyễn Đình Quân - 20521184, Nguyễn Hùng Phát - 22521074

December 10, 2025

LỜI CẢM ƠN

Trước tiên, em xin gửi lời cảm ơn chân thành và sâu sắc đến Ban Giám hiệu nhà trường và Khoa Khoa học Máy tính đã tạo điều kiện học tập và nghiên cứu thuận lợi trong suốt thời gian em theo học tại Trường Đại học Công nghệ Thông tin.

Em xin bày tỏ lòng biết ơn đặc biệt đến Thầy Đỗ Văn Tiến, đã trực tiếp giảng dạy và tận tình hướng dẫn em trong quá trình thực hiện đề tài khóa luận. Những định hướng, chỉ dẫn rõ ràng cùng sự hỗ trợ quý báu từ thầy đã là tiền đề quan trọng giúp em hoàn thành tốt công việc nghiên cứu và viết báo cáo đúng tiến độ. Em cũng xin cảm ơn thầy vì đã cung cấp tài liệu, giải đáp thắc mắc và luôn tạo môi trường học tập tích cực, hiệu quả.

Trong suốt quá trình thực hiện đề tài, em đã có cơ hội vận dụng những kiến thức nền tảng đã được học, đồng thời tích cực học hỏi, tìm tòi thêm các kiến thức mới. Đây là một trải nghiệm quý báu giúp em trưởng thành hơn trong tư duy và kỹ năng làm việc nghiên cứu.

Mặc dù đã nỗ lực hoàn thành đề tài với tinh thần nghiêm túc và cầu thị, nhưng do hạn chế về thời gian và kinh nghiệm, khóa luận không thể tránh khỏi những thiếu sót. Em rất mong nhận được sự thông cảm, góp ý chân thành từ các thầy cô để em có thể tiếp tục hoàn thiện và phát triển trong tương lai.

Em xin chân thành cảm ơn!

TÓM TẮT KHÓA LUẬN

aaaaa.....

Contents

List of Figures

List of Tables

Chapter 1

TỔNG QUAN

1.1 Đặt vấn đề

- **aa:** Ví dụ: ;
- **bb():** ;

Như vậy **đầu vào (Input)** của hệ thống bao gồm (xem thêm hình minh họa):

- **Tập dữ liệu video đầu vào:** Bao gồm các

Và **đầu ra (Output)** của hệ thống là:

- **Danh sách các:** Các đoạn video ngắn

Mặc dù đã có nhiều tiến bộ trong lĩnh vực, bài toán:

1. **Vấn đề 1 (vd1):**;
2. **Vấn đề 2 (vd2)::**;
3. **Vấn đề 3 (vd3)::**;

Từ những thách thức đã nêu, khóa luận này đặt ra mục tiêu phát triển một XXXXXXXXXXXX, có khả năng:

- XXX1;

- XXX2;
- XXXXX
- XXXXX

1.2 Mục tiêu và phạm vi

1.2.1 Mục tiêu

Trong bài khóa luận này, sinh viên đề ra các mục tiêu như sau:

- Tìm hiểu tổng quan về bài toán ;
- Thiết kế và triển khai một hệ thống ;
- Tìm hiểu việc cách

1.2.2 Phạm vi

Trong bài khóa luận này, nhóm sinh viên tập trung hoàn thành các công việc sau:

- Tìm hiểu các mô hình truy vấn ảnh hoặc video dựa trên mô tả ngôn ngữ tự nhiên, làm cơ sở cho việc xây dựng hệ thống baseline, trong đó có CLIP và các biến thể như BEiT-3;
- Phát triển hệ thống truy vấn video dựa trên mô tả văn bản sử dụng các mô hình đã khảo sát, đồng thời đánh giá hiệu quả về tài nguyên tính toán (bộ nhớ GPU, kích thước đặc trưng lưu trữ), thời gian tính toán và độ chính xác truy vấn;
- Tìm hiểu và đề xuất phương pháp kết hợp kết quả từ nhiều nguồn thông tin khác nhau (âm thanh, chữ trên video, đối tượng, ...) như Whisper cho bài toán trích xuất âm thanh (automatic speech recognition); DeepSolo và PARSeq cho bài toán nhận dạng chữ; và CO-DETR phát hiện đối tượng (object detection);

- Thiết kế mô-đun tái xếp hạng kết quả truy vấn dựa trên ngữ cảnh toàn cục, sử dụng mô hình ngôn ngữ đa phương thức (Multimodal LLM) để đánh giá lại độ liên quan giữa truy vấn và nội dung video ở cấp độ ngữ nghĩa sâu hơn;
- Xây dựng hệ thống truy vấn video tổng thể theo kiến trúc mô-đun, có khả năng tích hợp linh hoạt nhiều nguồn tín hiệu đầu vào và hỗ trợ các chiến lược truy vấn đa dạng.
- Thực hiện các thí nghiệm đánh giá hệ thống trên tập dữ liệu truy vấn thực tế, sử dụng các tiêu chí chuẩn như Reciprocal Rank (RR@K), xinfAP,... nhằm đo lường hiệu quả truy vấn trên nhiều khía cạnh.

1.3 Đóng góp của khóa luận

- XXXX1
- XXXX2
-

1.4 Cấu trúc khóa luận

Chương 1: Tổng quan bài toán.

Chương 2: Cơ sở lý thuyết và các nghiên cứu liên quan.

Chương 3: Một số phương pháp áp dụng để cải thiện hiệu quả cho bài toán truy vấn video.

Chương 4: Thực nghiệm và đánh giá.

Chương 5: Xây dựng ứng dụng minh họa.

Chương 6: Kết luận và hướng phát triển.

Chapter 2

CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN

2.1 Giới thiệu

2.1.1 Tổng quan và ý nghĩa thực tiễn của

2.1.2 Thách thức về tính

2.1.2.1 Bài toán)

Chapter 3

PHƯƠNG PHÁP

3.1 Hệ thống phát hiện và nhận dạng chữ trên biển hiệu

Chapter 4

THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1 Dữ liệu

4.2 Tiền xử lý

4.3 Tập câu truy vấn đánh giá

4.4 Độ đo đánh giá

4.5 Kết quả thực nghiệm

Chapter 5

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 Kết luận

.....:

- aaaa.

5.2 Hướng phát triển

Để khắc phục những hạn chế trên và nâng cao hơn nữa tính hiệu quả, tính khả dụng và tính mở rộng của hệ thống, các hướng phát triển trong tương lai được đề xuất như sau:

Tối ưu hóa khả năng mở rộng dữ liệu:

- aaa
- aaa

Tăng cường khả năng tương tác và thích ứng với người dùng:

- Thiết kế giao diện người dùng aaaaaaaaaaaaaa

Tích hợp truy vấn hình thức thoại (Spoken Query Integration): Phát triển hệ thống

.....