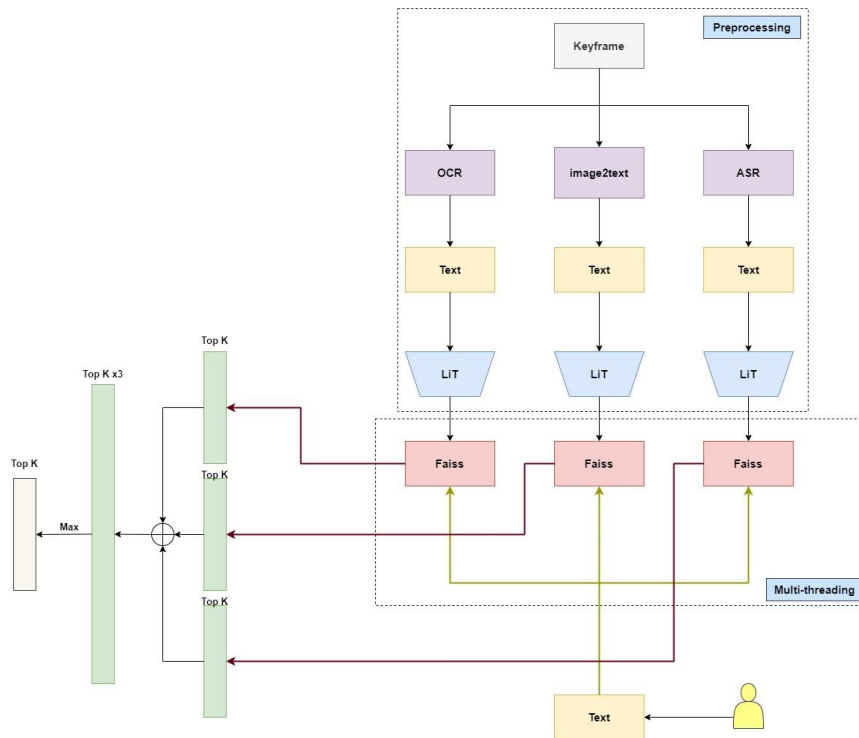
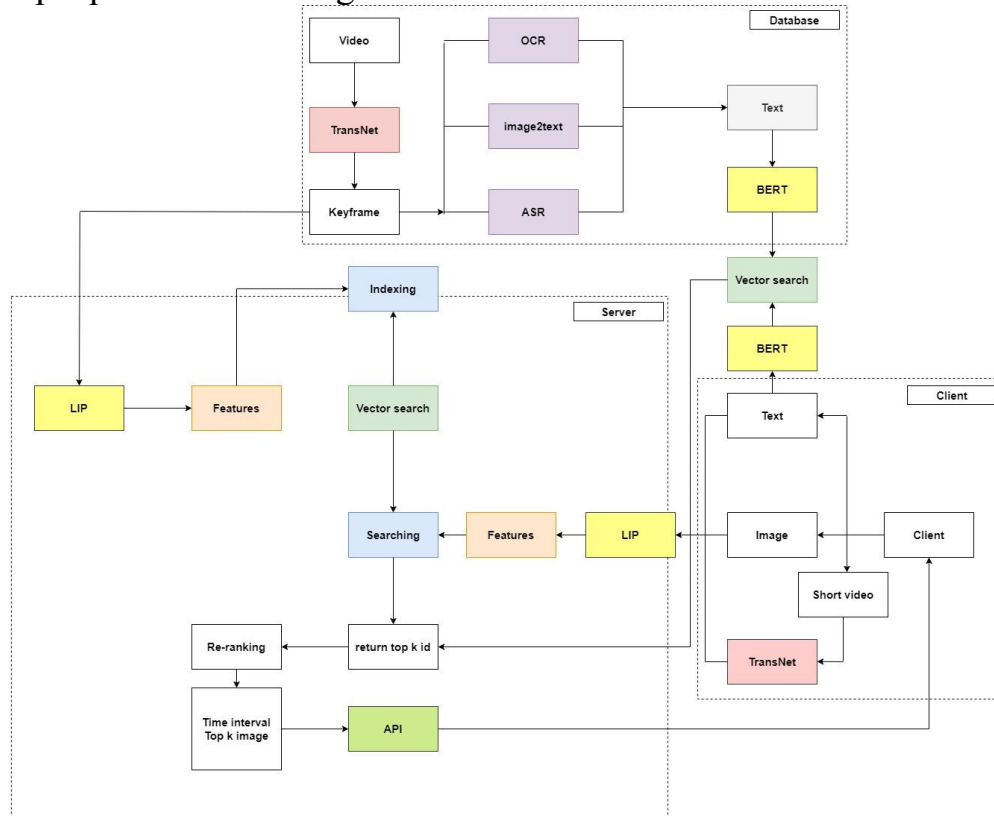


HCM AI Challenge 2023

UTE-AI-Fluc

● Mô tả giải pháp

Giải pháp của nhóm bao gồm các bước sau



Pipeline giải pháp

Hình 1.

● *Giai đoạn Database*

- Từ những tập dữ liệu mà ban tổ chức đã cung cấp, nhóm chúng em đã tách các video này thành các tập data chứa các keyframes bao gồm hình ảnh được cắt ra từ video cùng với ID tương ứng với từng hình ảnh. Nhóm đã sử dụng model TransNet để thực hiện điều này.
- Sau khi trích xuất được các keyframes, các keyframes này sẽ được đưa qua các model bao gồm OCR, ASR và Image2text để trích xuất thành những đoạn text, sau đó sử dụng Bert model để embedding đoạn text thành các vector search. Các vector search sẽ được đưa vào database để so sánh độ tương đồng cosine với các vector đã có sẵn.

● *Giai đoạn Server và Client*

- Sau khi có được bộ data chứa các keyframes, nhóm chúng em thực hiện chuyển các keyframes này thành các features thông qua sử dụng các model Blip, Clip và LiT (được gộp chung thành LIP). Sau đó đánh dấu Index cho từng các features và lưu dưới dạng file .bin.
- Khi người dùng thực hiện đưa vào các câu text, image hoặc các short video, các dữ liệu này sẽ được đưa vào khối model “LIP” để chuyển thành các features của input. Khi search, nó duyệt hết tính similarity với các features trong file .bin, từ đó chọn ra được top k feature có score cao sau đó qua Re-ranking để trả về các top k features có độ tương đồng cao với câu query.
- Các top k features đó sẽ được hiển thị lên cho hệ thống thông qua sử dụng API để người dùng lựa chọn.

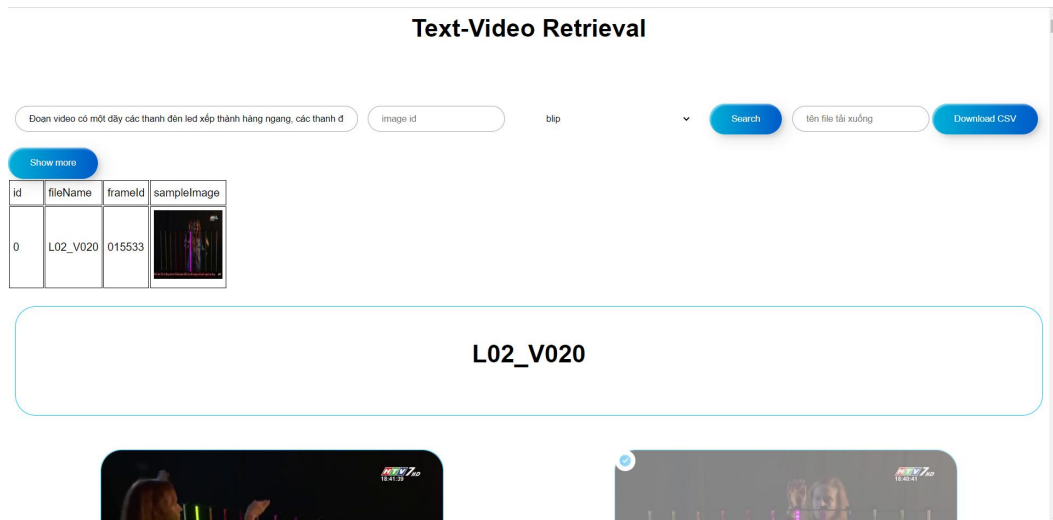
● *Giai đoạn Preprocessing và Multi-threading*

- Trong khối Preprocessing, bộ data chứa các keyframes sẽ được đưa qua các model OCR, ASR và Image2text trích xuất thành nhiều mỗi bộ text tương ứng theo mỗi model khác nhau, sau đó đưa các bộ text này qua LiT model để encode thành các vector, các output từ LiT model được đưa vào Faiss để similarity search.
- Trong khối Multi-threading, khi người dùng nhập đoạn text để search, các đoạn text này được đưa vào faiss để similarity search với các vector là các output của khối Preprocessing.
- Kết quả trả về là các bộ top k có scores cao tương ứng của từng model, thực hiện cộng toàn bộ các bộ top k đó thành bộ top k lớn hơn, sau đó tìm max của bộ top k này để tìm ra các features có scores cao đồng thời trả về cho người dùng các features có sự tương đồng cao với hình ảnh.

● *Xây dựng hệ thống*

Trang web hệ thống bao gồm các chức năng sau:

- Tìm kiếm bằng các câu text query.
- Tìm kiếm bằng các ID hình ảnh.
- Lựa chọn các chức năng khác nhau.
- Đặt tên cho file .csv.
- Cho phép tải về file csv chứa các keyframes.
- Button hiển thị thêm hình ảnh.



Hình 2: Trang web hệ thống

- + Tìm kiếm bằng các câu text query
Cho phép người dùng nhập các đoạn text mô tả hình ảnh vào ô “Query text” sau đó bấm search, kết quả nhận được là các keyframes tương ứng với đoạn text.
- + Tìm kiếm bằng các ID hình ảnh
Cho phép người dùng nhập các ID hình ảnh vào ô “Image ID” sau đó bấm search, kết quả nhận được là các keyframes tương tự như với ID hình ảnh đã search.
- + Lựa chọn các chế độ search khác nhau
Khi người dùng bấm vào ô “Please choose an option” sẽ hiển thị các chế độ search khác nhau, có 6 chế độ search bao gồm OCR, Image Captioning, Blip, Clip, LiT và ASR.
- + Đặt tên cho file .csv
Đây là chức năng cho phép người dùng có thể đặt tên cho các file .csv chứa các keyframes trước khi được tải xuống.
- + Cho phép tải về file .csv chứa các keyframes
Sau khi đã hoàn tất việc đặt tên file và chọn những hình ảnh tương thích, người dùng có thể nhấn vào button này để tải về file .csv chứa các keyframes đã chọn.
- + Button hiển thị thêm keyframes
Sau khi những keyframes đã hiển thị trên trang web hệ thống, nếu người dùng muốn tìm kiếm thêm những keyframes khác có thể chọn vào nút nhấn để show ra màn hình thêm những keyframes mới.

- **Kết quả của hệ thống**

Câu query: Đoạn video có một dãy các thanh đèn led xếp thành hàng ngang, các thanh đèn có màu khác nhau, được dựng đứng. Một người phụ nữ thực hiện động tác giống như gõ lên phía trên thanh đèn thì thanh đèn đó sẽ phát sáng.

Kết quả sau khi search:

Text-Video Retrieval

Đoạn video có một dãy các thanh đèn led xếp thành hàng ngang, các thanh đ

image id

blip


▼

Search



tên file tải xuống

Download CSV

Show more

id	fileName	frameId	sampleImage
0	L02_V020	015533	

L02_V020



Hình 3: Kết quả trả về sau khi tìm kiếm bằng câu query

Note: