

VIDEO CORPUS MOMENT RETRIEVAL

Phạm Huỳnh Long Vũ

Nguyễn Huy Phước

1 University of Information Technology

What ?

Đề tài nghiên cứu Video Corpus Moment Retrieval (VCMR): với một truy vấn ngắn ngủn tự nhiên và một kho video lớn, hệ thống cần:

- Truy hồi video liên quan nhất trong kho.
- Định vị chính xác đoạn thời gian $[t_{start}, t_{end}]$ khớp với mô tả.
- Xử lý cả truy vấn mô tả **một sự kiện** hoặc **chuỗi sự kiện có thứ tự**.

Baseline hướng tới **pipeline hai giai đoạn**: chi phí huấn luyện thấp, truy vấn nhanh, mở rộng tốt cho kho lớn.

Why ?

Tìm “đúng khoảnh khắc” trong video dài khó vì:

- Khoảnh khắc cần tìm thường **rất ngắn** so với toàn bộ video.
- Kho video **quy mô lớn** khiến việc quét toàn bộ rất tốn tài nguyên.
- Metadata dạng tiêu đề/mô tả thường **không đủ chi tiết** để định vị theo thời gian.

VCMR có ý nghĩa trong:

- tìm kiếm video học tập/giải trí,
- giám sát và truy vết sự kiện
- lập chỉ mục và quản lý dữ liệu đa phương tiện.

Overview



Figure 1 . Example

Description

1. Trích xuất keyframe

- Lấy keyframe với khoảng cách $\leq 1s$ để giảm bớt sót hành động ngắn.
- Loại khung hình lặp/na ná bằng đo độ tương đồng ảnh.
- Kết quả: bộ keyframe gọn, phục vụ truy vấn nhanh.

3. Giai đoạn 1 – Truy hỏi keyframe

- Mã hoá từng sự kiện và truy hỏi **top-k** keyframe trong không gian vector.
- Video là ứng viên nếu Giai đoạn 2 – Định vị khoảnh khắc (DP) có **ít nhất 1 keyframe phù hợp cho mỗi sự kiện**.
- Đầu ra: danh sách video ứng viên + keyframe ứng cử cho từng sự kiện.

5. Đánh giá

- **Recall@K**: chất lượng truy hỏi video.
- **Temporal IoU / IoU@threshold**: độ khớp thời gian định vị.
- **Query response time**: tốc độ phản hồi truy vấn.

2. Tách truy vấn thành chuỗi sự kiện

- Dùng LLM chuyển câu truy vấn thành danh sách sự kiện có thứ tự ($e_1 \rightarrow e_2 \rightarrow \dots \rightarrow e_n$).
- Hỗ trợ tốt cho truy vấn nhiều bước (multi-step).

4. Giai đoạn 2 – Định vị khoảnh khắc (DP)

- Áp đặt **ràng buộc thứ tự thời gian** của chuỗi sự kiện.
- Chọn chuỗi keyframe nhất quán → suy ra biên đoạn $[t_{start}, t_{end}]$.
- Xếp hạng theo **tổng/ghép** điểm tương đồng sự kiện-keyframe.

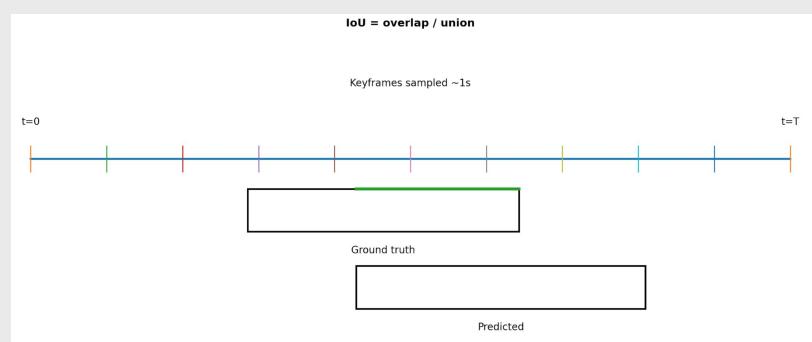


Figure 2 .Temporal IoU