

ĐỀ CƯƠNG NGHIÊN CỨU KHOA HỌC

TÊN ĐỀ TÀI: HƯỚNG TIẾP CẬN HỌC KHÔNG GIÁM SÁT CHO BÀI TOÁN TÓM TẮT VIDEO		
TÊN ĐỀ TÀI TIẾNG ANH: A NEW APPROACH TO UNSUPERVISED VIDEO SUMMARIZATION		
Cán bộ hướng dẫn: PGS.TS. Lê Đình Duy		
Thời gian thực hiện: Từ ngày .../2021 đến ngày .../2021.		
Sinh viên thực hiện:		
1. Trương Chí Diễm	19520464	Lớp: KHCL2019.1 Email: 19520464@gm.uit.edu.vn Điện thoại: 0832136510
2. Nguyễn Anh Dũng	19521394	Lớp: KHCL2019.1 Email: 19521394@gm.uit.edu.vn Điện thoại: 0855589818
3. Trần Hoàn Đức Duy	19521434	Lớp: KHCL2019.1 Email: 19521434@gm.uit.edu.vn Điện thoại: 0888193215

Nội dung đề tài:*(Mô tả chi tiết mục tiêu, phạm vi, đối tượng, phương pháp thực hiện, kết quả mong đợi của đề tài)*

Giới thiệu:

Ngày nay, số lượng video trên các nền tảng mạng xã hội như Youtube, Tiktok, Telegram,... gia tăng không ngừng do nhu cầu về giải trí cũng như phục vụ cho các mục đích khác. Cùng với sự phát triển của công nghệ, việc quay và đăng tải video ngày càng phổ biến. Theo trang [statista](#), cứ mỗi phút trôi qua, Youtube có khoảng 500 giờ video được đăng tải (số liệu năm 2020).

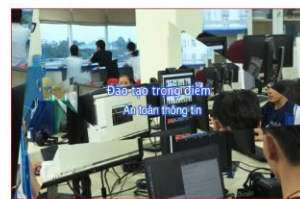
Việc xem hết tất cả video là bất khả thi. Xu hướng của người xem hiện nay là các video ngắn, dạng tóm tắt, nhưng vẫn giữ được thông tin nhiều nhất có thể. Tuy nhiên, việc sử dụng sức người để theo dõi hết video và trích lọc những phân đoạn tiêu biểu là rất tốn chi phí và không hiệu quả. Vì vậy bài toán tóm tắt video đang được quan tâm và thu hút nhiều nghiên cứu.

Các phương pháp học có giám sát như [1] [2] [3] đã được đề xuất nhưng vẫn còn một số điểm yếu cố hữu về việc tốn nhiều công sức gán nhãn và bộ dữ liệu còn thiếu sự đa dạng các lĩnh vực.

Trong đề tài này, chúng tôi tập trung nghiên cứu các phương pháp học không giám sát cho bài toán tóm tắt video để giải quyết các vấn đề được nêu ở trên. Các phương pháp đã được đề xuất như [4] [5] vẫn còn chưa đạt được hiệu năng cao. Vì vậy chúng tôi tìm một hướng tiếp cận mới, sử dụng học tăng cường để xây dựng mô hình tóm tắt video.

Input: Một video có độ dài từ 30 – 120 phút, độ phân giải không quá thấp.

Output: Video đã được tóm tắt, bao gồm các keyframes, độ phân giải giữ nguyên.



Mục tiêu:

- Nghiên cứu một cách áp dụng học tăng cường vào bài toán tóm tắt video, cho phép huấn luyện mô hình không cần nhãn.

Phạm vi:

- Vì lý do tài nguyên có giới hạn nên thời lượng video chỉ từ 30 – 120 phút, độ phân giải không quá thấp ($\geq 480p$).

Đối tượng:

- Tham khảo các phương pháp học tăng cường [6] [7] [8]. Đây là những phương pháp đang có kết quả cao và phù hợp với bài toán tóm tắt video.
- Reward function của [8] vẫn còn có thể sửa đổi và cải tiến thêm để tăng hiệu năng.

Nội dung:

- Sử dụng các bộ dữ liệu có sẵn để tiết kiệm tài nguyên con người và dễ dàng so sánh với các mô hình trước đó. Bộ dữ liệu TVSum [9] (thu thập từ Youtube gồm 50 videos). Bộ dữ liệu SumMe [10] (gồm 25 videos).
- Phân tích ưu, nhược điểm của các mô hình học tăng cường có sẵn. Kết hợp, sửa đổi, thử nghiệm trên các phiên bản khác nhau để tìm được mô hình có hiệu năng cao nhất.

- Tìm ra các reward functions mới, hợp lý hơn cho mô hình học tăng cường trong tóm tắt video.

Phương pháp:

- Chọn các Feature Extractors như [11] [12] [13].
- Tìm hiểu các tiêu chí về bài toán tóm tắt video. Sau đó tìm reward functions thích hợp.
- Tìm hiểu và lựa chọn độ đo hợp lý cho bài toán.
- Tiến hành huấn luyện, đánh giá và chọn mô hình tốt nhất.

Kết quả dự kiến:

- Bài báo khoa học về hướng tiếp cận mới theo phương pháp học tăng cường cho bài toán tóm tắt video.
- Đạt được kết quả đánh giá ngang bằng hoặc cao hơn các phương pháp hiện có.
- Công bố mã nguồn hoàn chỉnh và các trọng số mô hình đã được huấn luyện cho phương pháp đề xuất kèm hướng dẫn cài đặt và sử dụng.

Tài liệu tham khảo

- [1] GONG, Boqing, et al, "Diverse sequential subset selection for supervised video summarization," *Advances in neural information processing systems*, vol. 27, pp. 2069-2077, 2014.
- [2] ZHANG, Ke, et al, "Summary transfer: Exemplar-based subset selection for video summarization," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1059-1067, 2016.
- [3] ZHANG, Ke, et al, "Video Summarization with Long Short-term Memory," *European conference on computer vision*, pp. 766-782, Springer, Cham, 2016.
- [4] KHOSLA, Aditya, et al, "Large-scale video summarization using web-image priors," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2698-2705, 2013.
- [5] LI, Yingbo; Merialdo, Bernard, "Multi-video summarization based on video-mmrc," *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*, pp. 1-4, IEEE, 2010.
- [6] T. M. MOERLAND, J. BROEKENS and C. M. JONKER, "Model-based reinforcement learning: A survey," *arXiv preprint arXiv: 2006.16712*, 2020.
- [7] YANG, Huan, et al, "Unsupervised extraction of video highlights via robust recurrent auto-encoders," *Proceedings of the IEEE international conference on computer vision*, pp. 4633-4641, 2015.

- [8] ZHOU, Kaiyang; QIAO, Yu; XIANG, Tao, "Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.
- [9] SONG, Yale, et al, "TVSum: Summarizing web videos using titles," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5179-5187, 2015.
- [10] GYGLI, Michael, et al, "Creating summaries from user videos," *European conference on computer vision*, pp. 505-520, Springer, Cham, 2014.
- [11] SZEGEDY, Christian, et al, "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.
- [12] HE, Kaiming, et al, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
- [13] SIMONYAN, Karen; ZISSERMAN, Andrew, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

Kế hoạch thực hiện:(Mô tả tóm tắt kế hoạch làm việc và phân công công việc cho từng sinh viên tham gia)

Tuần 1 – 3: Khảo sát các phương pháp đã có, ghi chú, nhận xét kết quả.

Tuần 4 – 8: Lựa chọn các tiêu chí hợp lý để đánh giá mô hình và nghiên cứu, phát triển hàm reward cho bài toán.

Tuần 9 – 12: Thiết kế các mô hình khả thi, tiến hành cài đặt và rà soát lỗi.

Tuần 13 – 15: Tiến hành huấn luyện, đánh giá, ghi chú và chọn mô hình tốt nhất.

Tuần 16: Viết tài liệu và công bố kết quả.

Công cụ làm việc:

- Sử dụng công cụ Trello để phân công các nhiệm vụ theo từng spring.
- Dùng github để quản lý mã nguồn trong quá trình cài đặt và soát lỗi.
- Sử dụng LATEX để soạn thảo tài liệu được song song hóa.

<p>Xác nhận của CBHD (Ký tên và ghi rõ họ tên)</p> <p>Lê Đình Duy</p>	<p>TP. HCM, ngày....thángnăm 2021</p> <p>Sinh viên</p> <p>(Ký tên và ghi rõ họ tên)</p>
--	---