

# MỞ RỘNG BỐI CẢNH THỜI GIAN TRONG VIỆC HIỂU VIDEO BẰNG PHƯƠNG PHÁP NGÂN HÀNG ĐẶC TRƯNG DÀI HẠN

Nguyễn Duy Minh Nhật - 220202020

# Tóm tắt



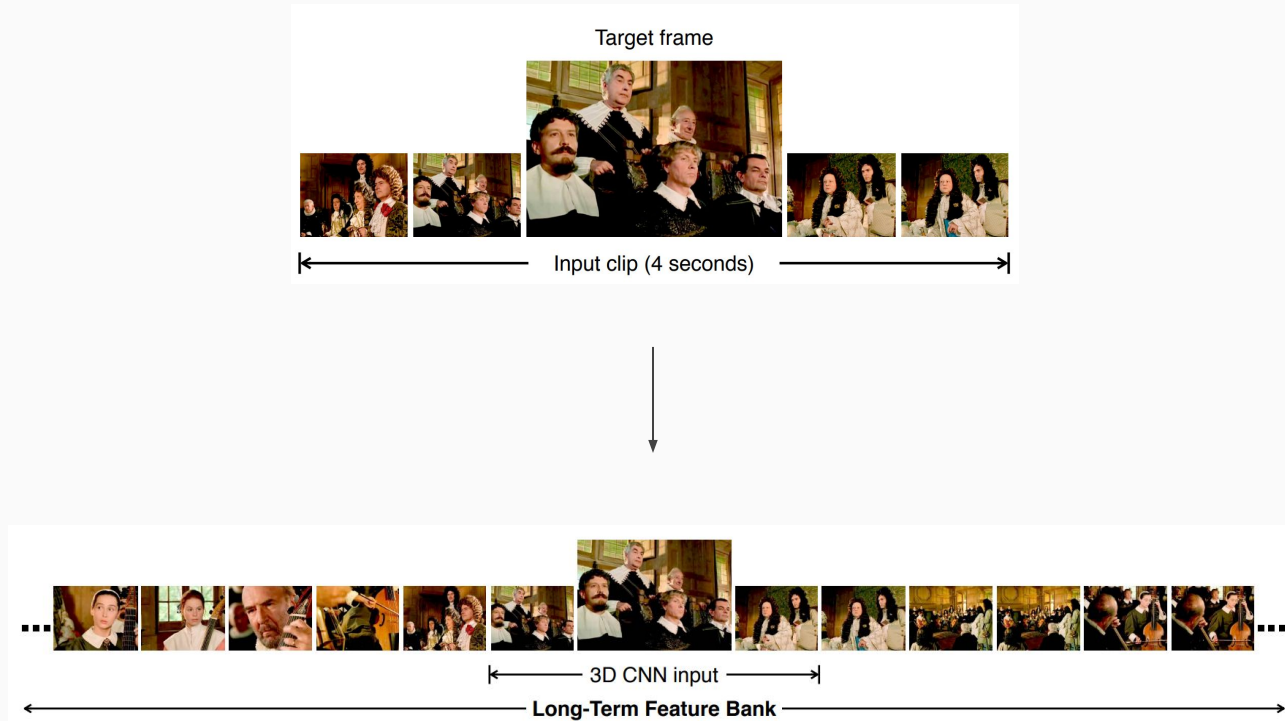
NGUYỄN DUY MINH NHẬT  
MSHV: 220202020

- Lớp: CS2205.APR2023
- Link Github:  
<https://github.com/nhatndmuit/CS2205.APR2023>
- Link YouTube video:  
<https://youtu.be/Bt3xZUkpUAg>

# Giới thiệu

- Hạn chế của các mô hình hiện tại: Các mô hình video hiện tại tập trung vào đoạn video ngắn, thiếu khả năng hiểu ngữ cảnh thời gian dài.
- Đề xuất một giải pháp mới nhằm cải thiện khả năng hiểu video bằng cách kết hợp thông tin bối cảnh dài hạn trong các mô hình video hiện có.

# Giới thiệu



# Mục tiêu

- Xây dựng một phương pháp để bổ sung khả năng hiểu video bằng cách sử dụng ngân hàng đặc trưng dài hạn.
- Tăng cường khả năng hiểu ngữ cảnh thời gian dài của các mô hình video state-of-the-art.
- Đánh giá mô hình được tăng cường trên các bộ dữ liệu video: AVA [4], EPIC-Kitchens [5] và Charades [6].

# Nội dung và Phương pháp

- Xây dựng ngân hàng đặc trưng dài hạn.
- Kết hợp ngân hàng đặc trưng dài hạn vào 3D CNN.
- Huấn luyện và thử nghiệm.
- Đánh giá hiệu suất.

# Kết quả dự kiến

- Nâng cao khả năng hiểu video bằng việc kết hợp ngân hàng đặc trưng dài hạn với mô hình 3D CNN.
- Đạt được kết quả tốt nhất trên các bộ dữ liệu video như AVA, EPIC-Kitchens và Charades.
- Khẳng định tính khả thi và hiệu quả của phương pháp đề xuất trong việc cải thiện việc hiểu video và liên kết sự kiện trong ngữ cảnh thời gian dài.

# Tài liệu tham khảo

- [1] João Carreira, Andrew Zisserman: Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. CVPR 2017: 4724-4733
- [2] Zhaofan Qiu, Ting Yao, Tao Mei: Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks. ICCV 2017: 5534-5542
- [3] Saining Xie, Chen Sun, Jonathan Huang, Zhuowen Tu, Kevin Murphy: Rethinking Spatiotemporal Feature Learning For Video Understanding. CoRR abs/1712.04851 (2017)
- [4] Chunhui Gu, Chen Sun, David A. Ross, Carl Vondrick, Caroline Pantofaru, Yeqing Li, Sudheendra Vijayanarasimhan, George Toderici, Susanna Ricco, Rahul Sukthankar, Cordelia Schmid, Jitendra Malik: AVA: A Video Dataset of Spatio-Temporally Localized Atomic Visual Actions. CVPR 2018: 6047-6056
- [5] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, Michael Wray: Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. CoRR abs/1804.02748 (2018)
- [6] Gunnar A. Sigurdsson, Gül Varol, Xiaolong Wang, Ali Farhadi, Ivan Laptev, Abhinav Gupta: Hollywood in Homes: Crowdsourcing Data Collection for Activity Understanding. ECCV (1) 2016: 510-526