

# **XEM XÉT VIỆC PHÂN LOẠI VIDEO BẰNG ZERO-SHOT: MÔ HÌNH HUẤN LUYỆN END-TO-END CHO CÁC ỨNG DỤNG THỰC TẾ**

**Vũ Quốc Huy - 19521642**

**Lê Văn Phước - 19522054**

# Tóm tắt

- Lớp: CS519.M11
- Link Github của nhóm:  
<https://github.com/levanphuoc-dev/CS519.M11>
- Link YouTube video:
- Nhóm PHP:



Lê Văn Phước - 19522054



Vũ Quốc Huy - 19521642

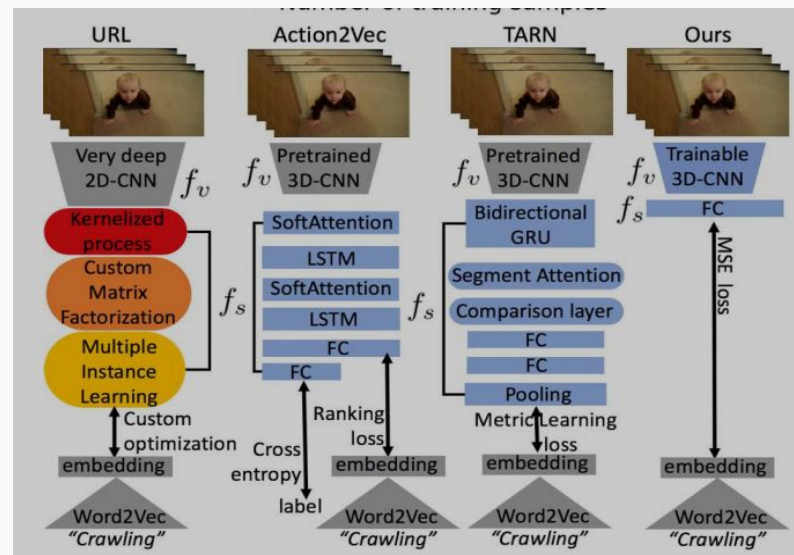
# Giới thiệu

Những mô hình hiện đại:

- Mô hình phức tạp
- Khó tái sản xuất
- Hiệu suất kém

Mô hình Zero-shot learning:

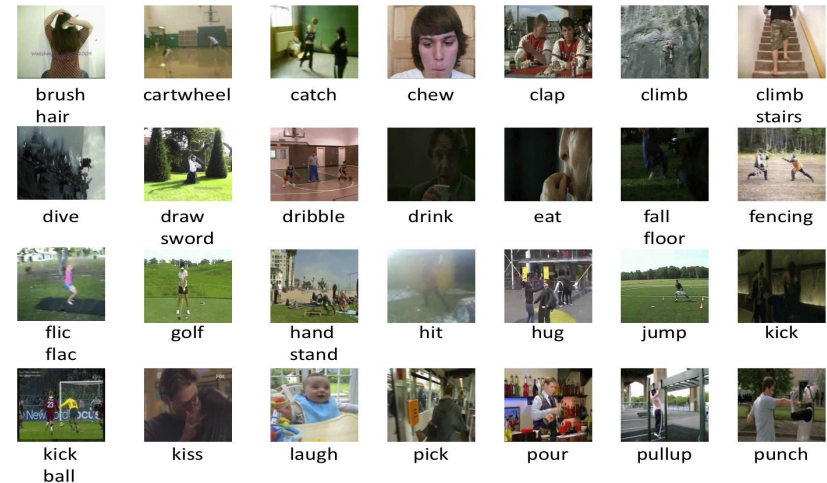
- Đơn giản hơn
- Dễ tái sản xuất
- Hiệu suất tốt hơn



# Giới thiệu dataset

UCF101: có 101 lớp hành động. 13320 videos nguồn từ Youtube. Các lớp hành động chủ yếu xoay quanh hành động thể thao

HMDB51: có 51 lớp hành động. 6767 videos. Các lớp hành động chủ yếu xoay quanh hành động thể thao và các hoạt động ngày thường của con người.



# Mục tiêu

- Đề xuất được mô hình Zero-shot Learning trong việc nhận dạng hành động trong video.
- Đề xuất một phương thức đào tạo (end-to-end training) và đánh giá mới cho mô hình Zero-shot Learning để thực thi và cài đặt cho thực tế.
- Phân tích chuyên sâu mô hình để tìm ra đặc điểm của một bộ dữ liệu tốt cho việc huấn luyện mô hình Zero-shot learning.

# Ứng dụng

Dùng để phân loại video dựa trên hành động, rồi đề xuất cho người xem có nhu cầu xem hoặc tìm kiếm video dựa theo hành động.

Ví dụ: một người dùng đăng tải video thể thao về trượt tuyết, mô hình sẽ có thể tự động gán nhãn cho video, và có thể đề xuất cho người khác một cách dễ dàng khi họ tìm kiếm video có nội dung liên quan đến hành động đó.

# Nội dung

- Đề xuất một phương thức đào tạo và đánh giá mô hình Zero-shot Learning để thực thi và cài đặt cho thực tế.
- So sánh với các mô hình hiện đại hiện có như URL, Action2Vec, TARN.
- Phân tích chuyên sâu mô hình để tìm ra những đặc điểm của một bộ dữ liệu tốt để huấn luyện mô hình Zero-shot Learning.
- Kết luận.

# Phương pháp

- Định nghĩa mô hình Zero-shot learning cho bối cảnh phân loại video.
- Sử dụng phương thức huấn luyện End-to-end cho Zero-shot Learning.
- Thiết kế một phương pháp tăng tập dữ liệu để tạo ra các video dùng để huấn luyện từ các ảnh tĩnh.
- Huấn luyện mô hình trên bộ dữ liệu.
- Tìm hiểu cách đánh giá mô hình.
- So sánh mô hình Zero-shot Learning dùng phương thức huấn luyện End-to-end với những mô hình hiện đại hiện có.
- Phân tích chuyên sâu mô hình để tìm ra đặc điểm của một bộ dữ liệu tốt cho Zero-shot Learning.
- Kết luận



# Kết quả dự kiến

- Tạo ra được mô hình Zero-shot Learning dùng phương thức huấn luyện End-to-end sử dụng để phân loại hành động trên Video. Và mô hình này có hiệu suất vượt trội hơn so với các phương pháp hiện có.
- Tạo ra một kỹ thuật tiền đào tạo (pretraining) mới cho Zero-shot Learning để giúp nhận dạng hành động video tốt hơn.
- Tìm ra được dưới điều kiện nào thì mô hình đào tạo End-to-end có hiệu suất vượt trội, để có thể áp dụng vào thực tế một cách tốt nhất.

# Tài liệu tham khảo

- Biagio Brattoli, Joseph Tighe, Fedor Zhdanov, Pietro Perona, Krzysztof Chalupka. Rethinking Zero-Shot Video Classification: End-to-End Training for Realistic Applications. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 4613-4623.