

PHÁT HIỆN GIAN LẬN TRONG THI TRỰC TUYỂN THÔNG QUA PHÂN TÍCH VIDEO WEBCAM VÀ ÂM THANH SỬ DỤNG MÔ HÌNH TRANSFORMER ĐA MODAL

Học viên: Nguyễn Lê Vinh - 240101084
GVHD: PGS. TS Lê Đình Duy

Tóm tắt

- Lớp: **CS2205.FEB2025**
- Link Github của nhóm:
<https://github.com/NLeVinh/CS2205.FEB2025>
- Link YouTube video:
<https://youtu.be/5ho-zSubVTA>
- Họ và tên: **Nguyễn Lê Vinh - 240101084**

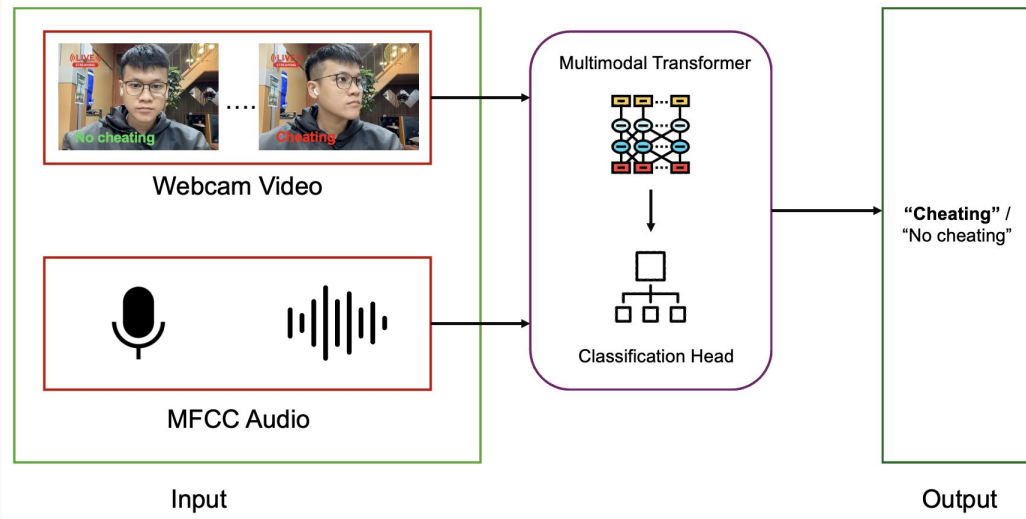


Giới thiệu

- Gian lận thi cử trực tuyến ngày càng phổ biến trong bối cảnh học tập và kiểm tra từ xa
- Các hành vi gian lận thường tinh vi, khó phát hiện: thì thầm, ra hiệu, nhìn ra ngoài, sử dụng tai nghe,...
- Một số nghiên cứu [1, 2, 3] đã áp dụng mô hình CNN, LSTM để phân tích riêng lẻ video hoặc âm thanh.
- Gần đây, một mô hình CNN-LSTM kết hợp đa modal cho thấy kết quả khả quan hơn [4], nhưng vẫn còn hạn chế:
 - LSTM khó học quan hệ dài hạn
 - Chưa khai thác hiệu quả tương quan giữa các modal
- Mô hình đề xuất: Vision Transformer (video) + Transformer Encoder (âm thanh) bằng cross-attention

Giới thiệu

- Input:
 - Chuỗi khung hình từ video webcam
 - Dữ liệu âm thanh
- Output: Nhãn dự đoán hành vi "Cheating" hoặc "No cheating"



Hình 1: Sơ đồ mô hình phát hiện gian lận thi trực tuyến từ video webcam và âm thanh

Mục tiêu

- Nghiên cứu và xây dựng mô hình Transformer đa modal kết hợp dữ liệu video webcam và âm thanh môi trường để phát hiện hành vi gian lận trong thi trực tuyến.
- Thực nghiệm và so sánh hiệu quả giữa các mô hình thông qua các độ đo Accuracy, Precision, Recall, F1-score:
 - Đơn modal (ViT, Audio Transformer)
 - CNN-LSTM
 - Transformer đa modal đề xuất
- Xây dựng ứng dụng prototype mô phỏng hệ thống giám sát thi trực tuyến, cho phép người dùng tải video, âm thanh và tự động đánh giá hành vi của thí sinh.

Nội dung và Phương pháp

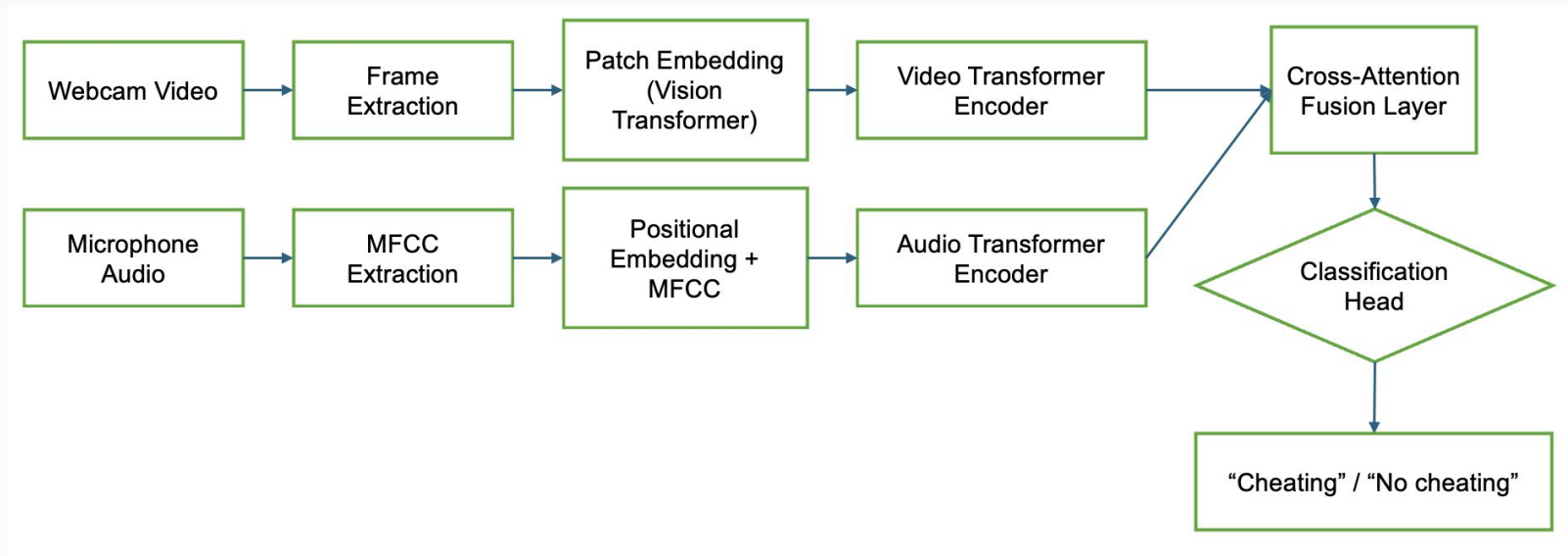
- Nội dung
 - Tìm hiểu các hành vi gian lận phổ biến trong thi trực tuyến.
 - Nghiên cứu mô hình ViT cho video, Transformer encoder cho MFCC âm thanh.
 - Tìm hiểu cơ chế cross-attention trong mô hình Transformer đa modal.
 - Đọc hiểu và sử dụng bộ dữ liệu công khai MSU Online Exam Proctoring Dataset¹ (video + âm thanh + nhãn hành vi)
 - Huấn luyện và so sánh các mô hình đơn modal (ViT, Audio Transformer), mô hình tích hợp CNN-LSTM baseline và mô hình Transformer đa modal đề xuất.
 - Xây dựng ứng dụng prototype minh họa hệ thống giám sát.

¹ Sumit Saboo, S. Lamba, Nidhi Sharma: MSU Online Exam Proctoring Dataset. Mendeley Data, V1 (2021).
<https://www.kaggle.com/datasets/sumitsaboo/msu-online-exam-proctoring-dataset>

Nội dung và Phương pháp

- Phương pháp
 - Tìm hiểu hành vi gian lận từ tài liệu, báo cáo và video thực tế.
 - Nghiên cứu các mô hình từ các bài báo, hội nghị về học sâu đa modal [5, 6, 7].
 - Tiền xử lý dữ liệu MSU Dataset: trích khung hình video (OpenCV), trích xuất đặc trưng MFCC từ âm thanh (Librosa).
 - Huấn luyện và đánh giá mô hình với Accuracy, Precision, Recall, F1-score [8]
 - Triển khai ứng dụng web: tải video - âm thanh, phân tích và đưa ra kết quả.

Nội dung và Phương pháp



Hình 2: Pipeline mô hình Transformer đa modal phát hiện gian lận thi trực tuyến.

Kết quả dự kiến

- Báo cáo kết quả thực nghiệm các mô hình phát hiện gian lận thi trực tuyến: đơn modal, CNN-LSTM, và hướng tiếp cận Transformer đa modal.
- So sánh hiệu quả giữa các mô hình qua các độ đo: Accuracy, Precision, Recall, F1-score.
- Xây dựng ứng dụng web prototype phát hiện gian lận qua video và âm thanh, hiển thị cảnh báo hành vi gian lận.
- Thử nghiệm ứng dụng trong môi trường lớp học mô phỏng để tiếp tục thu thập dữ liệu và cải thiện mô hình.

Tài liệu tham khảo

- [1]. Mahesh Navale, Aryan Jadhav, Mahesh Kadam, Shalmali Karandikar, Siddhi Kate: "A Computer Vision-Based Solution for Exam Cheating Detection." *International Journal of Innovative Research in Information Security*, Vol. 3(5), pp. 20–25 (2021).
- [2]. Kamran Shaukat, Shuai Luo, Vijay Varadharajan, Imran A. Hameed, Sheng Chen, Dake Liu, Jinhai Li: "Student Cheating Detection in Higher Education by Implementing Machine Learning Techniques." *Sensors*, Vol. 23(8), Article 4149 (2023).
- [3]. Zhiwei Hu, Yifan Jing, Guoping Wu, Hongyi Wang: "Multi-Perspective Adaptive Paperless Examination Cheating Detection System Based on Image Recognition." *Applied Sciences*, Vol. 14(10), Article 4048 (2024).
- [4]. Sumit Lamba, Nidhi Sharma: "Deep Learning-Based Multimodal Cheating Detection in Online Proctored Exams." *Journal of Electrical Systems*, Vol. 20(3), pp. 7375–7383 (2024).
- [5]. Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, Cordelia Schmid: "ViViT: A Video Vision Transformer." *CoRR*, abs/2103.15691 (2021).
- [6]. C. S. Sonali, Chinmayi B S, Ahana Balasubramanian: "Transformer-based Sequence Labeling for Audio Classification based on MFCCs." *CoRR*, abs/2305.00417 (2023).
- [7]. Yun Liu, Zheng Li, Tianyang Xu, Yang Chen, Yi Zhang: "Multimodal Transformer Using Cross-Channel Attention for Object Detection." *CoRR*, abs/2310.13876 (2023).
- [8]. David M. W. Powers: "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation." *Journal of Machine Learning Technologies*, Vol. 2(1), pp. 37–63 (2011).