


# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://www.youtube.com/watch?v=of-8oqIs-mk>
- Link slides (dạng .pdf đặt trên Github của nhóm):  
[CS519.O11/DEVELOPMENT\\_AND\\_OPTIMIZATION\\_OF\\_REAL\\_TIME\\_GESTURE\\_BEHAVIOR\\_RECOGNITION\\_ON\\_MOBILE\\_DEVICES.pdf at main · BeQuang/CS519.O11 \(github.com\)](https://github.com/BeQuang/CS519.O11/DEVELOPMENT_AND_OPTIMIZATION_OF_REAL_TIME_GESTURE_BEHAVIOR_RECOGNITION_ON_MOBILE_DEVICES.pdf)
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

<ul style="list-style-type: none"><li>● Họ và Tên: Bùi Thành Quang</li><li>● MSSV: 21522507</li></ul> 	<ul style="list-style-type: none"><li>● Lớp CS519.O11</li><li>● Tự đánh giá (tổng điểm kết quả môn): 9/10</li><li>● Số buổi vắng: 0</li><li>● Số câu hỏi QT cá nhân: 10</li><li>● Số câu hỏi QT của cả nhóm: 1</li><li>● Link Github:</li><li>● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none"><li>○ Lên ý tưởng cho đề án</li><li>○ Tìm hiểu nội dung kiến thức</li><li>○ Viết báo cáo đề án, viết slide và poster</li><li>○ Làm video Youtube</li><li>○ Chỉnh sửa, thiết kế nội dung báo cáo, Slide và Poster</li><li>○ Làm video Youtube</li></ul></li></ul>
---	---



# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

PHÁT TRIỂN VÀ TỐI ƯU HÓA NHẬN DIỆN CỬ CHỈ/ HÀNH VI NGƯỜI TRONG THỜI GIAN THỰC TRÊN THIẾT BỊ DI ĐỘNG

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

DEVELOPMENT AND OPTIMIZATION OF REAL-TIME GESTURE/BEHAVIOR RECOGNITION ON MOBILE DEVICES

## TÓM TẮT (Tối đa 400 từ)

Real-time Image-Matching trên thiết bị di động là một nhiệm vụ đầy thách thức do tài nguyên tính toán hạn chế và yêu cầu thời gian đáp ứng nghiêm ngặt. Nghiên cứu này đề xuất kết hợp học máy vào quá trình kết hợp hình ảnh thời gian thực trên thiết bị di động để cải thiện hiệu suất và chính xác. Các thuật toán học máy sẽ được áp dụng trong giai đoạn trích xuất đặc trưng, kết hợp và xác minh hình học, nhằm tăng cường khả năng nhận dạng và khả năng chống nhiễu của hệ thống. Ngoài ra, cũng sẽ tiếp tục nghiên cứu các phương pháp tối ưu phần cứng để tận dụng khả năng tính toán của thiết bị di động. Kết quả nghiên cứu sẽ cung cấp một giải pháp toàn diện cho kết hợp hình ảnh thời gian thực trên thiết bị di động, kết hợp cả các thuật toán học máy và tối ưu phần cứng

## GIỚI THIỆU (Tối đa 1 trang A4)

Việc áp dụng phương pháp Image-Matching trên thời gian thực đã được thực hiện từ rất lâu và đã dần trở nên phổ biến trong thời gian trở lại đây vì tính ứng dụng cao, áp dụng được cho nhiều bài toán cụ thể trong cuộc sống. Nhưng việc phát triển một mô hình để trở nên tiện dụng, dễ tiếp cận và đạt hiệu quả cao trên các thiết bị di động là một bài toán khó và đang ngày càng hoàn thiện hơn.

Sử dụng thư viện TensorFlow Lite[1] giúp cho việc áp dụng các model máy học lên các thiết bị di động được thực hiện dễ dàng hơn. Từ đó, ta có thể thực hiện

huấn luyện mô hình so khớp gương mặt người phát hiện được và phát hiện hành vi xấu được trích xuất từ 1 realtime camera và báo tín hiệu lại cho người sử dụng thiết bị di động.

Trong đề tài này, tôi sẽ thực hiện tìm hiểu về việc áp dụng mô hình nhận diện hành vi xấu trên camera thời gian thực, từ đó tìm ra các ưu điểm/nhược điểm nhằm hướng tới việc áp dụng hiệu quả và tối ưu hóa mô hình, ứng dụng trên các thiết bị cầm tay. Việc này sẽ áp dụng một số mô hình máy học, thư viện đã được đào tạo sẵn có như là “Nhận diện hành vi con người”, thư viện “TensorFlow Lite/Model Maker”[2], thư viện “OpenCV”[3],...

Input: 1 video của 1 người đang có một hành vi bất kỳ.

Output: Nhận diện được danh tính của người có trong video và hành vi của người đó có trong video, nếu hành vi đó là sai thì sẽ báo lại tín hiệu.

## **MỤC TIÊU**

*(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)*

- Phát triển mô hình học máy để tăng cường khả năng nhận dạng và khả năng chống nhiễu của hệ thống kết hợp hình ảnh trên thiết bị di động.
- Xây dựng được hệ thống camera tích hợp mô hình máy học để nhanh chóng phát hiện người có hành vi xấu nhất về thiết bị di động cho người dùng

## **NỘI DUNG VÀ PHƯƠNG PHÁP**

*(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)*

### **Nội du**

- Thực hiện xây dựng bộ dữ liệu và huấn luyện ra một mô hình giúp phát hiện cử chỉ, hành vi của người dựa trên một thiết bị thu hình ảnh, video thời

gian thực(camera, máy ảnh, điện thoại,...)

- Chuyển đổi và tích hợp mô hình trên một thiết bị di động dưới dạng một ứng dụng.
- Rút ra các ưu, nhược điểm của mô hình và các hướng phát triển, cải thiện mô hình để tối ưu hơn trên thiết bị di động.

## **Phương pháp**

- Thu thập dữ liệu: Thu thập một tập dữ liệu lớn chứa hình ảnh hoặc video chứa các cử chỉ và hành vi người mà bạn muốn nhận diện. Đảm bảo rằng các cử chỉ và hành vi đã được gắn nhãn đúng cho từng khung hình hoặc video.
- Tiền xử lý dữ liệu: Tiền xử lý dữ liệu để chuẩn bị cho việc huấn luyện mô hình. Các bước tiền xử lý có thể bao gồm áp dụng thư viện OpenCV để cắt ảnh hoặc video thành các khung hình, thay đổi kích thước ảnh để đảm bảo đồng nhất kích thước đầu vào cho mô hình, chuẩn hóa dữ liệu để đưa các giá trị pixel về phạm vi thích hợp, và chuyển đổi định dạng dữ liệu thành định dạng phù hợp cho mô hình máy học.
- Xây dựng mô hình máy học: Chọn một kiến trúc mô hình phù hợp cho nhiệm vụ nhận diện cử chỉ và hành vi người. Các kiến trúc phổ biến cho xử lý dữ liệu hình ảnh và video bao gồm Convolutional Neural Networks (CNNs)[4] và Recurrent Neural Networks (RNNs)[5]. Bạn cũng có thể sử dụng các mô hình pre-trained như I3D[6], TSN[7] hoặc C3D[8] và điều chỉnh lại cho nhiệm vụ của mình bằng cách huấn luyện tiếp trên dữ liệu của bạn.
- Huấn luyện mô hình: Sử dụng tập dữ liệu huấn luyện đã gắn nhãn, huấn luyện mô hình bằng cách tối ưu hóa hàm mất mát thông qua việc điều chỉnh các trọng số của mô hình. Điều này có thể được thực hiện bằng cách sử

dụng các thuật toán tối ưu như stochastic gradient descent (SGD)[9] hoặc Adam[10]. Đảm bảo chia tập dữ liệu thành tập huấn luyện và tập kiểm tra để đánh giá hiệu suất của mô hình.

- **Đánh giá và điều chỉnh mô hình:** Đánh giá mô hình trên tập kiểm tra để đo lường độ chính xác và hiệu suất của nó. Nếu kết quả không đạt yêu cầu, bạn có thể điều chỉnh mô hình bằng cách thay đổi kiến trúc, thêm hoặc bớt các lớp, hoặc tăng kích thước tập dữ liệu huấn luyện.
- **Triển khai mô hình trên thiết bị thời gian thực:** Sau khi mô hình đã được huấn luyện và đánh giá, chuyển đổi mô hình sang định dạng phù hợp để triển khai trên thiết bị thời gian thực như TensorFlow Lite. Điều này giúp thu nhỏ kích thước mô hình và tối ưu hóa hiệu suất cho thiết bị có tài nguyên hạn chế.
- **Tích hợp và triển khai mô hình:** Sử dụng TensorFlow Lite hoặc các công cụ phù hợp khác để tích hợp mô hình vào ứng dụng trên thiết bị thời gian thực. Thiết lập luồng dữ liệu đầu vào từ cảm biến hoặc camera và sử dụng mô hình để phát hiện cử chỉ và hành vi người trong thời gian thực. Kết quả nhận diện có thể được sử dụng để thực hiện các hành động phụ hợp, như đưa ra cảnh báo, ghi lại thông tin hoặc tương tác với người dùng.

## KẾT QUẢ MONG ĐỢI

*(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)*

- *Các thuật toán trích xuất và kết hợp đặc trưng hiệu quả được thiết kế đặc biệt cho việc kết hợp hình ảnh thời gian thực trên thiết bị di động.*
- *Các phương pháp xác minh hình học được tối ưu hóa phù hợp cho nền tảng di động, nâng cao tính ổn định của việc kết hợp hình ảnh.*
- *Các kỹ thuật tối ưu phần cứng và tối ưu hóa để tận dụng tài nguyên tính toán có sẵn trên thiết bị di động.*

- Một khung công việc kết hợp hình ảnh thời gian thực cho thiết bị di động, đảm bảo hiệu suất kết hợp hình ảnh chính xác và nhanh chóng.
- Kết quả đánh giá so sánh cho thấy tính hiệu quả và tính hiệu quả của các thuật toán và tối ưu hóa đề xuất.

### **TÀI LIỆU THAM KHẢO** (Định dạng DBLP)

- [1] TensorFlow: A system for large-scale machine learning - Zhifeng Chen, Google Inc.
- [2] Machine Learning for Mobile Developers: Tensorflow Lite Framework - Avid Farhoodfar Sofia University, IEEE Consumer Electronics Society SCV
- [3] OpenCV for Computer Vision Applications - Naveenkumar Mahamkali, (SRM University-AP), Vadivel Ayyasamy (SRM University-AP), Proceedings of National Conference on Big Data and Cloud Computing (NCBDC'15)
- [4] An Introduction to Convolutional Neural Networks - Keiron O'Shea, Ryan Nash
- [5] Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network - Alex Sherstinsky
- [6] Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset - Joao Carreira, Andrew Zisserman
- [7] Temporal Segment Networks for Action Recognition in Videos - Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, Luc Van Gool
- [8] Learning Spatiotemporal Features with 3D Convolutional Networks - Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri

### **KẾ HOẠCH THỰC HIỆN:**

+ Tuần 1-4: Tìm hiểu các mô hình nhận diện cử chỉ/hành vi người.

Kết quả dự tính:

- Hiểu cấu trúc, thuật toán của các phương pháp
- Tài liệu chi tiết về các đặc điểm và ưu/nhược điểm của các mô hình

+ Tuần 4-12: Xây dựng mô hình với các ưu điểm của các mô hình và đánh giá mô hình.

Kết quả dự tính:

- Bảng kết quả đánh giá của mô hình mới trên bộ dữ liệu tự thu thập

+ Tuần 10-16: Xây dựng ứng dụng bằng TensorFlow Lite Model Maker

Kết quả dự tính:

- Tối ưu được mô hình trên 1 thiết bị di động cụ thể (điện thoại di động) và sử dụng thông qua kết nối với 1 thiết bị camera bằng kết nối Wifi
- Báo tín hiệu về thiết bị di động nếu trên camera xuất hiện hành vi được cho là xấu.