

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**



BÁO CÁO ĐỒ ÁN CUỐI KÌ

Môn học: Máy học

Đề tài: Phát hiện cảnh bạo lực trong video

Lớp: CS114.k21

Giảng viên hướng dẫn: PGS.Ts Lê Đình Duy

ThS: Phạm Nguyễn Trường An

Thành phố Hồ Chí Minh, Ngày 14 Tháng 8 Năm 2020

MỤC LỤC

I. GIỚI THIỆU NHÓM	3
1.1 Tên nhóm: Vô Danh	3
1.2 Thành viên:	3
1.3 Phân công công việc:	3
1.4 Đánh giá tổng quan nhóm:	4
1.5 Đánh giá nhận xét thành viên:	4
II. GIỚI THIỆU ĐỀ TÀI	4
2.1 Bài toàn	4
2.2 Mục tiêu đề tài	5
2.3 Ứng dụng	5
III. CÁC BƯỚC THỰC HIỆN	6
3.1 Thu thập dữ liệu	6
3.2 Xử lý dữ liệu	6
3.3 Phân chia Data	6
IV. TRÍCH XUẤT ĐẶC TRƯNG	7
4.1 Lấy mẫu thừa thớt	7
4.2 Phát hiện vùng chuyển động	8
4.3 Trích xuất lấy đặc trưng	9
V. HUẤN LUYỆN MODEL	10
VI. KẾT QUẢ	11
VII. TỔNG KẾT	12
VIII. TÀI LIỆU THAM KHẢO	12

I. GIỚI THIỆU NHÓM

1.1 Tên nhóm: Vô Danh

1.2 Thành viên:

+ Trịnh Quang Trường – 18520393 – CS114.k21

+ Tăng Năng Chung – 18520536 – CS114.K21

+ Phan Lê Phú – 18521247 – CS114.k21

1.3 Phân công công việc:

Tên thành viên	Công Việc	Mức độ hoàn thành
Trịnh Quang Trường	-Chỉ đạo nhóm -Tìm ý tưởng -Tách video thành clip -Viết chương trình -Kiểm tra chạy chương trình -Làm báo cáo, soạn PowerPoint	- Hoàn thành xuất sắc (100/100)
Tăng Năng Chung	-Tìm ý tưởng -Tìm và Download video -Lọc Clip từ việc cắt video -Kiểm tra, chạy thử chương trình -Làm báo cáo, soạn PowerPoint	-Hoàn thành tốt (90/100)
Phan Lê Phú	-Tìm ý tưởng -Lọc Clip từ việc cắt video -Viết chương trình -Kiểm tra, chạy thử chương trình -Làm báo cáo, soạn PowerPoint	-Hoàn thành tốt (90/100)

1.4 Đánh giá tổng quan nhóm:

- + Các thành viên đều tham gia tốt hoạt động nhóm.
- + Nhóm hòa đồng thân thiện, mọi thành viên đều hòa thuận và không có mâu thuẫn.
- + Đôi khi nhóm cũng xảy ra một chút lười biếng nhưng cũng không ảnh hưởng đến công việc.

1.5 Đánh giá nhận xét thành viên:

- + Trịnh Quang Trường (Trưởng nhóm): là người tiên phong dẫn dắt các thành viên của nhóm, luôn giao nhiệm vụ cho các thành viên khác. Thúc đẩy các thành viên làm việc cũng như tạo ra những buổi họp mặt thành viên để trao đổi. Công việc của mình thì hoàn thành tốt không xảy ra sai phạm.
- + Tăng Năng Chung(thành viên): Là người được trưởng nhóm giao nhiệm vụ, chấp hành kỉ luật của nhóm, luôn tham gia đầy đủ những buổi gặp mặt cũng như trao đổi với các thành viên của nhóm. Về công việc được giao cũng hoàn thành tốt, đạt tiến trình mà nhóm đề ra nhưng đôi khi cần sự giúp đỡ của các thành viên khác.
- + Phan Lê Phú(Thành viên): Là người được trưởng nhóm giao nhiệm vụ, chấp hành kỉ luật của nhóm, luôn tham gia đầy đủ những buổi gặp mặt cũng như trao đổi với các thành viên của nhóm. Về công việc được giao cũng hoàn thành tốt, đạt tiến trình nhóm đề ra.

II. GIỚI THIỆU ĐỀ TÀI

2.1 Bài toàn

Hiện nay thường xuyên xảy ra các vụ mâu thuẫn ẩu đả, nhiều vụ còn xảy ra đâm chém dẫn đến chết người. Nguyên nhân xảy đến các vụ mâu thuẫn thường là cãi

nhau từ lâu, thù oán, hay uất ức nhiều khi có những vụ xảy ra do rượu bia, chất kích thích...

Các cuộc đụng độ có thể được người dân phát hiện báo với chính quyền còn có những vụ xảy ra rất nặng nhưng lại không hề hay biết vì xảy ra âm thầm hoặc người dân sợ liên lụy không dám báo cáo.

Ngoài ra trên mạng xã hội ngày này có quá nhiều nội dung độc hại, đánh nhau, ẩu đả cũng là một trong số đó làm ảnh hưởng đến tư tưởng người dùng đặc biệt là trẻ nhỏ.

Vì những lí do trên nhóm chúng em quyết định xây dựng một mô hình giúp nhận diện bạo lực.

Input: một video bất kì.

Output: phát hiện những cảnh đánh nhau trong video(nếu có).



Hình 1 Cảnh bình thường



Hình 2 Cảnh đánh nhau.

2.2 Mục tiêu đề tài

Xây dựng được một mô hình giúp phân loại video và lọc các video bạo lực ảnh hưởng trên mạng xã hội.

Phát hiện và thông báo đến các cơ quan chức năng những vụ đánh nhau trên đường phố thông qua camera giám sát.

Là tiền đề, cơ sở giúp phát triển các ứng dụng cao hơn.

2.3 Ứng dụng

Xây dựng mô hình giúp phân loại những video bạo lực độc hại trên mạng xã hội tránh sự tiếp thu sai ảnh hưởng đến trẻ em.

Dựa vào sự quan sát của các camera đường phố tạo ra một ứng dụng giúp nhận diện phát hiện và thông báo đến các cơ quan chính quyền khi phát hiện đánh nhau xảy ra trên đường phố. Nhằm giảm thiểu tối đa tác hại của vụ ẩu đả gây ra.

III. CÁC BƯỚC THỰC HIỆN

3.1 Thu thập dữ liệu

Dữ liệu nhóm tự thu thập bằng việc download các video có cảnh đánh nhau được camera quan sát ghi lại trên mạng xã hội như facebook, youtube... Với tổng số video thu thập được khoảng 200 video

3.2 Xử lý dữ liệu

Dữ liệu video khi mới download về chưa thể sử dụng mà phải trải qua quá trình xử lý.

Video gốc được cắt ra thành những clip nhỏ được giới hạn khoảng 90 khung hình với thời lượng 3s.

Tiến hành lọc clip sau khi cắt, lọc ra những clip chưa toàn cảnh đánh nhau và những clip không chứa cảnh đánh nhau, những clip hỗn loạn, không rõ, hay dư thừa thì xóa bỏ.

Dữ liệu thu được sau khi lọc được khoảng 817 clip trong đó:

- + 405 clip toàn cảnh đánh nhau.
- + 412 clip không chứa cảnh đánh nhau.

Dữ liệu thu được sẽ được tiến hành gán nhãn mới 0 cho những clip đánh nhau và 1 cho những clip không chứa đánh nhau.

3.3 Phân chia Data

Mục đích chia data: thuận tiện cho quá trình train model, tạo ra ba bộ data để sau quá trình training trên mỗi bộ có thể so sánh kết quả và rút ra được kết luận.

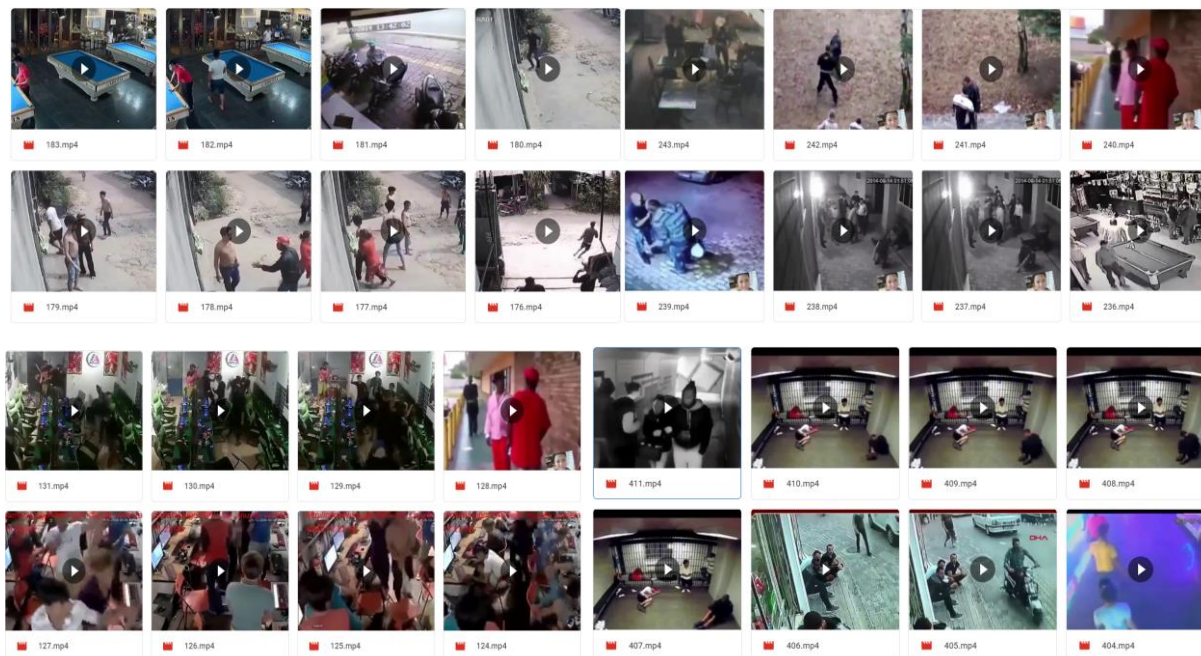
Bộ dữ liệu sau khi thu được tiếp tục chia làm 3 bộ data:

- + Bộ 1: gồm 400 clip(200 bạo lực và 200 không bạo lực)

+ Bộ 2: gồm 417 clip(205 bạo lực và 217 không bạo lực)

+ Bộ 3: toàn bộ 817 clip(405 bạo lực và 417 không bạo lực)

Một số hình ảnh mô tả bộ dữ liệu:



IV. TRÍCH XUẤT ĐẶC TRƯNG

4.1 Lấy mẫu thưa thớt

Do giữa các frame liên tục không có sự thay đổi rõ rệt, việc lấy hết các frame này là không cần thiết. Vì vậy, chúng em quyết định chia nhỏ clip 3s thành 15 đoạn bằng nhau, lấy frame đầu tiên của mỗi đoạn.

Với bộ dữ liệu sau bước tiền xử lý, lấy được tổng cộng:

+ Có cảnh đánh nhau: hơn 4000 frame

+ Không có cảnh đánh nhau: hơn 4100 frame



Hình 3 Frame trước



Hình 4 Frame sau cách 5 frame trước

Vì video thu thập có nhiều kích cỡ frame khác nhau, cần resize về cùng một kích thước. Kích thước được chọn là 160x120.



Hình 5 Ảnh trước khi resize

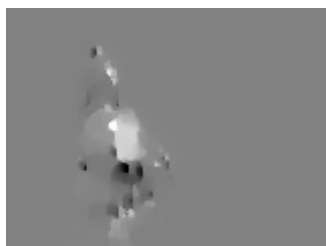


Hình 6 Ảnh sau khi resize và đổi sang định dạng Grayscale

4.2 Phát hiện vùng chuyển động

Đối với video có cảnh đánh nhau nhưng camera được đặt cố định, việc phát hiện ra các đối tượng tương đối dễ dàng. Tuy nhiên, với bộ dữ liệu được thu thập từ nhiều nguồn trên Internet nên dẫn đến việc xuất hiện các video bị giật khung hình. Những video này không thể sử dụng background subtraction để tách đối tượng chuyển động ra được. Sau khi tìm hiểu trong một số thuật toán [1], chúng em nhận thấy việc trích xuất Optical Flow khả thi trong việc thực hiện nhiệm vụ này.

Với độ chính xác cao, thuật toán TVL1 được dùng để trích xuất Optical Flow. Dùng 2 frame liên tiếp đã chuyển về định dạng grayscale, thu được ảnh optical flow theo hai hướng x và y. Dùng công thức [2] để chuyển về ảnh biểu diễn cường độ chuyển động (Mag).



Hình 7 Optical Flow theo hướng x

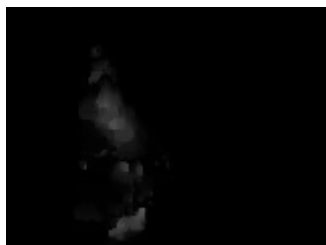


Hình 8 Optical Flow theo hướng y



Hình 9 Cường độ chuyển động (Mag)

Sau khi có được Mag, dùng thuật toán Canny [3] để phát hiện cạnh. Sau đó dùng phương thức Closing có trong thư viện OpenCV để nối liền các đường nét và lấp đầy các vùng có đường nét bao phủ. Những vùng đó cũng chính là vùng chuyển động.



Hình 10 Cường độ chuyển động (Mag)



Hình 11 Phát hiện cạnh bằng Canny



Hình 12 Lấp đầy cạnh bằng Opening

Tuy nhiên có một số lỗ nhỏ nằm bên trong vùng chuyển động. Ta lấp đầy những lỗ nhỏ bằng sử dụng hàm drawContours trong OpenCV.



Hình 13 Tồn tại các lỗ nhỏ



Hình 14 Sau khi lấp đầy bằng drawContours

Cuối cùng dùng hàm Opening để xóa đi những đường nét còn sót lại, ta thu được ảnh nhị phân biểu diễn vùng chuyển động.

4.3 Trích xuất lấy đặc trưng

Coi vùng chuyển động như là một vùng quan tâm (ROI), tiến hành gắn nó vào trong frame grayscale đang xét.



Hình 15 Frame đang xét



Hình 16 ROI



Hình 17 Sau khi gắn ROI

Trích xuất HOG từ ảnh vừa tìm được, ta thu được vector đặc trưng.

V. HUẤN LUYỆN MODEL

Theo như phân phân chia dữ liệu (data) ở trên ta có:

+ Bộ dữ liệu 1: Chứa 3607 examples, 1749 examples được gán nhãn có bạo lực, 1858 examples được gán nhãn không bạo lực và mỗi example là một vector đặc trưng được trích xuất bởi HOG có độ dài 54288.

+ Bộ dữ liệu 2: Chứa 3584 examples, 1837 examples được gán nhãn có bạo lực, 1737 examples được gán nhãn không bạo lực và mỗi example là một vector đặc trưng được trích xuất bởi HOG có độ dài 54288.

+ Bộ dữ liệu 3 : Được kết hợp bởi bộ dữ liệu 1 và 2. Chứa 7191 examples, 3586 examples được gán nhãn có bạo lực, 3605 examples được gán nhãn không bạo lực và mỗi example là một vector đặc trưng được trích xuất bởi HOG có độ dài 54288.

Vì đề một bài toán binary classification. Đối với bài toán này đã có rất nhiều thuật toán máy học để huấn luyện model do đó nhóm chúng em quyết định kiểm thử với nhiều thuật toán. Sau đó em sẽ chọn model cho kết quả tốt nhất dựa trên các thông số precision, recall, fscore.

Dưới đây là các thuật toán nhóm em đã chọn để huấn luyện model.

- + Logistic Regression
- + GaussianNB
- + Random Forest Classifier
- + SVM

VI. KẾT QUẢ

Kết quả thu được sau khi train model với từng bộ dữ liệu.

+ Bộ dữ liệu 1

	precision	recall	fscore	support
LogisticRegression	0.886810	0.887131	0.886419	None
GaussianNB	0.759103	0.759381	0.758957	None
RandomForestClassifier	0.734528	0.727175	0.726810	None
SVM	0.876573	0.876918	0.876663	None

+ Bộ dữ liệu 2

	precision	recall	fscore	support
LogisticRegression	0.872105	0.873515	0.872603	None
GaussianNB	0.675747	0.674303	0.674724	None
RandomForestClassifier	0.701454	0.695463	0.696201	None
SVM	0.884854	0.885173	0.885006	None

+ Kết hợp hai bộ dữ liệu 1 và 2

	precision	recall	fscore	support
LogisticRegression	0.754144	0.753318	0.752505	None
GaussianNB	0.610604	0.608937	0.608099	None
RandomForestClassifier	0.631559	0.627982	0.626364	None
SVM	0.786569	0.786585	0.786577	None

Nhận xét : Khi bộ dữ liệu càng lớn thì kết quả chính xác của từng model đều giảm. Kết quả của mỗi model còn thấp chưa thực sự được cao. Điều này có thể do bước tiền xử lý dữ liệu chưa thực sự được tốt, phương pháp trích xuất đặc trưng (HOG) có thể chưa phải là phương pháp tối ưu cho bài toán này, chưa thực hiện tinh chỉnh hyperparameter cho từng model, bộ dữ liệu dùng để huấn luyện chưa thực sự chuẩn và còn nhiều nhiễu, Mặc dù vậy kết quả như vậy vẫn có thể cho một hiệu quả đối với sản phẩm demo nên chúng em vẫn sẽ sử dụng model train được. Với từng bộ dữ liệu khác nhau ta nhận thấy model sử dụng thuật toán svm luôn cho kết quả cao hơn các model còn lại. Vì vậy chúng em quyết định sử dụng model đó áp dụng cho demo kết quả.

VII. TỔNG KẾT

Tổng kết: Kết quả demo của sản phẩm vẫn còn cho kết quả hiển thị vẫn chưa được đúng chính xác như mong muốn. Sau khi kiểm thử trên nhiều video kết quả thu về chỉ ở mức tạm chấp nhận được. Để khắc phục những nhược điểm tồn tại. Nhóm chúng em đề xuất cần cải thiện hơn nữa ở các bước tiền xử lý dữ liệu, trích xuất đặc trưng, train model và thu thập nhiều dữ liệu hơn. Sau đây là những phần chúng em đạt được và những khó khăn chúng em gặp phải khi xử lý bài toán này.

+ Đạt được:

- Hiểu và thực hiện được quy trình khi xử lý các vấn đề liên quan tới máy học.
- Biết một số kỹ thuật rút trích đặc trưng đơn giản
- Kỹ thuật xử lý ảnh đơn giản.
- Tạo một sản phẩm demo đơn giản

+ Khó khăn:

- Thu thập và gán nhãn dữ liệu.
- Tiền xử lý hình ảnh.
- Xử lý dữ liệu lớn.
- Tài nguyên hạn chế.

VIII. TÀI LIỆU THAM KHẢO

1. Tutorial optical flow (OpenCv) ([Link](#))
2. Zhou P, Ding Q, Luo H, Hou X (2018) Violence detection in surveillance video using low-level features. ([Link](#))
3. Canny Edge Detection (OpenCv) ([Link](#))