

THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
(ví dụ: <https://www.youtube.com/watch?v=AWq7uw-36Ng>)
- Link slides (dạng .pdf đặt trên Github của nhóm):
(ví dụ: <https://github.com/mynameuit/CS2205.xxx/TenDeTai.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*
- *Lớp Cao học, mỗi nhóm một thành viên*

- Họ và Tên: Nguyễn Thanh Hùng
- MSSV: 250101025
- Lớp: CS2205.CH201
- Tự đánh giá (điểm tổng kết môn): 8.5/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 2
- Số câu hỏi QT của cả nhóm: 0
- Link Github:
<https://github.com/mynameuit/CS2205.xxx/>



ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

BÀI TOÁN THEO DÕI ĐA ĐỐI TƯỢNG TRONG VIDEO VÀ ỨNG DỤNG VÀO
BÀI TOÁN ĐÉM XE(Car) GIAO THÔNG

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

MULTI-OBJECT TRACKING IN VIDEO AND ITS APPLICATION TO VEHICLE
(Car) COUNTING

TÓM TẮT (*Tối đa 400 từ*)

Theo dõi đa đối tượng trong video (Multi-Object Tracking – MOT) là một bài toán nền tảng trong thị giác máy tính, có vai trò quan trọng trong các ứng dụng giám sát giao thông. Mục tiêu của bài toán MOT là duy trì định danh nhất quán cho nhiều đối tượng xuất hiện liên tiếp trong chuỗi video, trong bối cảnh dữ liệu thường bị ảnh hưởng bởi nhiễu, che khuất và sự thay đổi theo thời gian.

Các nghiên cứu gần đây cho thấy hiệu quả theo dõi không chỉ phụ thuộc vào mô hình phát hiện đối tượng, mà còn chịu ảnh hưởng lớn từ cách liên kết các kết quả phát hiện giữa các khung hình. ByteTrack (ECCV 2022) là một phương pháp tiêu biểu theo hướng này, khi khai thác cả các bounding box có độ tin cậy thấp nhằm cải thiện khả năng theo dõi trong các kịch bản phức tạp.

Trong phạm vi môn học, đề tài này không hướng đến việc đề xuất mô hình học sâu mới, mà tập trung nghiên cứu bài toán theo dõi đa đối tượng dựa trên ByteTrack và phân tích khả năng ứng dụng vào bài toán đếm xe trong video giao thông. Đầu vào là dữ liệu video giám sát, đầu ra là số lượng xe đi qua một vị trí xác định. Thông qua pipeline phát hiện – theo dõi – đếm, đề tài làm rõ vai trò của yếu tố thời gian trong dữ liệu video và khả năng mở rộng cho các bài toán phân tích giao thông thông minh.

GIỚI THIỆU (*Tối đa 1 trang A4*)

Trong các hệ thống giám sát hiện nay, video là nguồn dữ liệu phổ biến để phân tích các hoạt động diễn ra trong môi trường thực, đặc biệt trong các bài toán liên quan đến giao thông. Tuy nhiên, việc khai thác hiệu quả dữ liệu video không đơn giản, bởi thông tin không chỉ nằm ở từng khung hình riêng lẻ mà còn ở mối quan hệ giữa các khung hình theo thời gian.

Một trong những bài toán cốt lõi khi xử lý video là theo dõi đa đối tượng, với mục tiêu duy trì định danh nhất quán cho nhiều đối tượng xuất hiện liên tiếp trong chuỗi khung hình. Bài toán này đóng vai trò nền tảng cho nhiều ứng dụng thực tế. Chẳng hạn, trong bài toán đếm xe giao thông, nếu mỗi phương tiện không được theo dõi liên tục theo thời gian, hệ thống rất dễ gặp các lỗi như đếm trùng cùng một xe hoặc bỏ sót xe khi xảy ra che khuất.

Từ thực tế đó, việc nghiên cứu bài toán theo dõi đa đối tượng được xem là bước cần thiết trước khi xây dựng các hệ thống đếm dựa trên video. Thay vì tiếp cận trực tiếp bài toán đếm, đề tài này lựa chọn đi từ bài toán nền tảng nhằm làm rõ vai trò của theo dõi đối tượng trong việc đảm bảo tính ổn định và chính xác của kết quả đếm.

Trong số các công trình nghiên cứu gần đây, ByteTrack (ECCV 2022) được lựa chọn làm cơ sở tham khảo chính cho đề tài. Lý do lựa chọn là phương pháp này không chỉ đạt hiệu quả cao trong các bộ dữ liệu chuẩn mà còn có cách tiếp cận đơn giản, phù hợp với việc phân tích và áp dụng trong phạm vi một học phần. Dựa trên nghiên cứu đó, đề tài tập trung phân tích khả năng ứng dụng bài toán theo dõi đa đối tượng vào việc đếm xe trong video giao thông, thay vì đề xuất một phương pháp học sâu mới.

MỤC TIÊU (*Viết trong vòng 3 mục tiêu*)

Đề tài này được thực hiện với ba mục tiêu chính.

Thứ nhất, đề tài nhằm làm rõ bài toán theo dõi đa đối tượng trong video dưới góc nhìn phương pháp luận nghiên cứu khoa học. Thay vì tiếp cận bài toán ở mức triển khai kỹ thuật, mục tiêu ở đây là hiểu bản chất của bài toán, các giả định đặt ra, cũng như vai trò của theo dõi đối tượng trong các hệ thống phân tích video.

Thứ hai, đề tài tập trung nghiên cứu phương pháp ByteTrack và phân tích lý do vì sao phương pháp này có thể cải thiện khả năng theo dõi trong các tình huống phức tạp như mật độ đối tượng cao hoặc xảy ra che khuất. Việc nghiên cứu ByteTrack được thực hiện với mục đích rút ra các ý tưởng phương pháp, không nhằm đề xuất hay cải tiến thuật toán mới.

Thứ ba, dựa trên bài toán theo dõi đa đối tượng, đề tài hướng tới việc xây dựng một quy trình ứng dụng bài toán này vào việc đếm xe trong video giao thông. Mục tiêu cuối cùng là đánh giá tính khả thi của cách tiếp cận này và làm rõ mối liên hệ giữa bài toán theo dõi đa đối tượng và bài toán đếm dựa trên video.

NỘI DUNG VÀ PHƯƠNG PHÁP

Đề tài được tiếp cận theo hướng phân tích và kế thừa từ các nghiên cứu trước, trong đó ByteTrack được lựa chọn làm phương pháp tham khảo chính. Việc lựa chọn này xuất phát từ thực tế rằng ByteTrack là một phương pháp tiêu biểu, có mã nguồn công khai và đã được kiểm chứng trên nhiều bộ dữ liệu chuẩn, phù hợp cho việc nghiên

cứu trong phạm vi học phần.

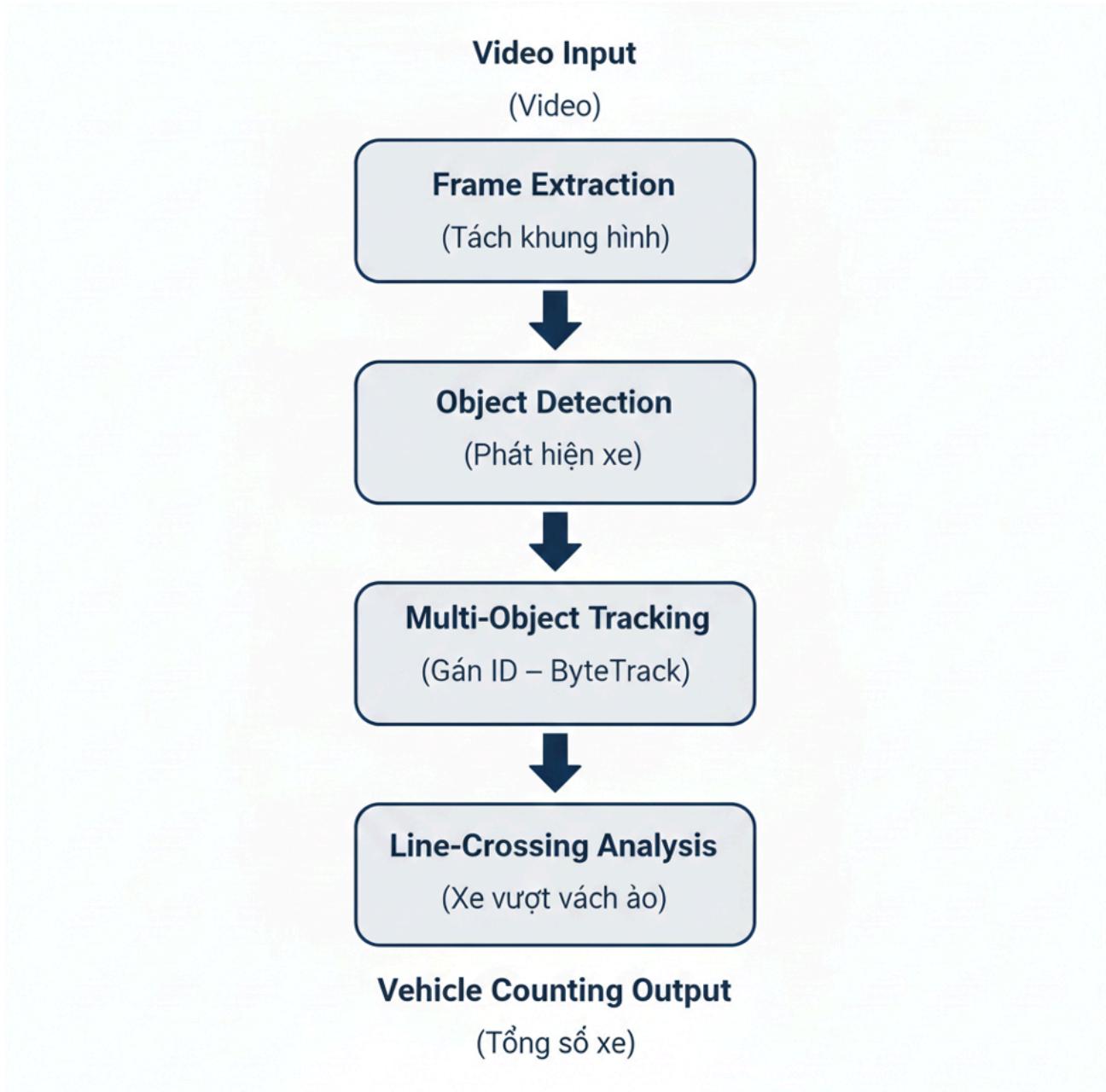
Trước hết, dữ liệu đầu vào của bài toán là các đoạn video giao thông thu thập từ các nguồn công khai. Các video này được xử lý theo từng khung hình liên tiếp nhằm phục vụ cho các bước xử lý tiếp theo. Trên mỗi khung hình, một mô hình phát hiện đối tượng được sử dụng để xác định vị trí các xe xuất hiện trong cảnh. Trong đề tài này, mô hình phát hiện được xem như một thành phần hỗ trợ, không phải trọng tâm nghiên cứu.

Tiếp theo, các kết quả phát hiện từ từng khung hình được đưa vào thuật toán ByteTrack để thực hiện liên kết đối tượng giữa các khung hình liên tiếp. Thay vì chỉ sử dụng các bounding box có độ tin cậy cao, ByteTrack khai thác thêm các phát hiện

có độ tin cậy thấp nhằm duy trì định danh đối tượng trong các tình huống khó, chẳng hạn khi xe bị che khuất tạm thời hoặc xuất hiện trong thời gian ngắn.

Sau khi mỗi xe được gán một định danh nhất quán theo thời gian, bài toán đếm được thực hiện bằng cách xác định một vạch cắt ngang ảo trong khung hình. Khi một đối tượng có định danh duy nhất vượt qua vạch này, hệ thống ghi nhận một sự kiện đếm. Cách tiếp cận này giúp giảm thiểu các lỗi đếm trùng và làm rõ vai trò của theo dõi đa đối tượng trong bài toán đếm xe.

Toàn bộ quy trình trên được xây dựng nhằm thể hiện tư duy phân rã bài toán, trừu tượng hóa và thiết kế quy trình xử lý – những yếu tố cốt lõi trong phương pháp luận nghiên cứu khoa học.



KẾT QUẢ MONG ĐỢI

Thông qua việc nghiên cứu bài toán theo dõi đa đối tượng trong video và kế thừa phương pháp ByteTrack, để tài kỹ vọng xây dựng được một quy trình đếm xe từ video giao thông có cấu trúc rõ ràng và logic. Quy trình này giúp làm rõ mối quan hệ giữa bài toán theo dõi đối tượng và bài toán đếm, thay vì xem đếm xe như một tác vụ độc lập.

Kết quả mong đợi đầu tiên của đề tài là việc làm rõ vai trò của theo dõi đa đối tượng trong việc đảm bảo tính ổn định của hệ thống đếm. Cụ thể, khi mỗi xe được gắn một định danh nhất quán theo thời gian, hệ thống đếm dự kiến sẽ giảm đáng kể các lỗi phổ biến như đếm trùng một phương tiện hoặc bỏ sót phương tiện khi xảy ra che khuất tạm thời.

Bên cạnh đó, để tài kỹ vọng cho thấy rằng việc khai thác các phát hiện có độ tin cậy thấp, như cách tiếp cận của ByteTrack, có thể mang lại lợi ích trong các kịch bản giao thông phức tạp, nơi các phương tiện thường xuyên xuất hiện và biến mất trong thời gian ngắn. Điều này giúp hệ thống đếm hoạt động ổn định hơn so với các phương pháp chỉ dựa trên phát hiện đối tượng ở từng khung hình riêng lẻ.

TÀI LIỆU THAM KHẢO (*Định dạng DBLP*)

[1] Yifu Zhang, Peize Sun, Yi Jiang, et al. ByteTrack: Multi-Object Tracking by Associating Every Detection Box. ECCV, 2022.