

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học:

**CS519 - PHƯƠNG PHÁP LUẬN
NGHIÊN CỨU KHOA HỌC**

Lớp học:

CS519.N11

Giảng viên:

PGS.TS. LÊ ĐÌNH DUY


Thời gian:



09/2022 - 02/2023

----- *Trang này cố tình để trống* -----

THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
<https://www.youtube.com/watch?v=Gv8zh8FDGH4>
- Link slides (dạng .pdf đặt trên Github của nhóm):
<https://github.com/Banhkun/CS519.N11/blob/main/CS519.N11.DeCuong.FinalReport.AIO.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">● Họ và Tên: Lê Việt Thịnh● MSSV: 20520781 	<ul style="list-style-type: none">● Lớp: CS519.N11● Tự đánh giá (điểm tổng kết môn): 9/10● Số buổi vắng: 2● Số câu hỏi QT cá nhân: 8● Số câu hỏi QT của cả nhóm: 3● Link Github: https://github.com/Banhkun/CS519.N11● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none">○ Lên ý tưởng sử dụng và kết hợp đặc trưng.○ Viết phần nội dung và phương pháp.○ Kiểm tra các lỗi văn bản, lỗi trình bày.
<ul style="list-style-type: none">● Họ và Tên: Đinh Nhật Minh● MSSV: 20521597	<ul style="list-style-type: none">● Lớp: CS519.N11● Tự đánh giá (điểm tổng kết môn): 9/10● Số buổi vắng: 8● Số câu hỏi QT cá nhân: 0● Số câu hỏi QT của cả nhóm: 0

	<ul style="list-style-type: none"> ● Link Github: https://github.com/Banhkun/CS519.N11 ● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: <ul style="list-style-type: none"> ○ Viết phần tóm tắt, giới thiệu, mục tiêu và kết quả mong đợi ○ Thống kê lại tài liệu tham khảo ○ Làm trang Github, viết documentation.
<ul style="list-style-type: none"> ● Họ và Tên: Dương Thành Bảo Khanh ● MSSV: 20521444 	<ul style="list-style-type: none"> ● Lớp: CS519.N11 ● Tự đánh giá (điểm tổng kết môn): 8.5/10 ● Số buổi vắng: 1 ● Số câu hỏi QT cá nhân: 5 ● Số câu hỏi QT của cả nhóm: 3 ● Link Github: https://github.com/Banhkun/CS519.N11 ● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: <ul style="list-style-type: none"> ○ Làm bài trình chiếu ○ Trình bày về đồ án

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

CẢI TIẾN NHẬN DIỆN HÀNH ĐỘNG TRONG BÀI TOÁN PHÁT HIỆN TƯƠNG TÁC NGƯỜI VẬT SỬ DỤNG ĐA ĐẶC TRƯNG CHO KIẾN TRÚC TRANSFORMER

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

IMPROVING ACTION RECOGNITION IN HUMAN-OBJECT INTERACTION DETECTION PROBLEM USING MULTI-FEATURE FUSION FOR TRANSFORMER ARCHITECTURE

TÓM TẮT *(Tối đa 400 từ)*

Bài toán phát hiện tương tác người và vật đã và đang thu hút sự quan tâm của cộng đồng thị giác máy tính trong những năm gần đây. Vấn đề do bài toán đặt ra là làm sao để ghép cặp giữa con người và đối tượng, sau đó mô tả hành động tương tác của họ (con người). Theo thời gian, đã có nhiều phương pháp được đưa ra để giải quyết vấn đề này, tuy nhiên các phương pháp này chỉ tập trung chủ yếu vào các đặc trưng hình ảnh mà bỏ qua các thông tin có giá trị khác. Mục tiêu của đề tài là tìm hiểu và khảo sát các nghiên cứu liên quan trên bài toán phát hiện tương tác người vật; khảo sát hiệu quả của các đặc trưng thị giác, ngữ nghĩa và không gian và sử dụng các mô hình dựa trên transformer để huấn luyện. Qua các thử nghiệm, đề tài sẽ đề xuất một cách ứng dụng các đặc trưng đạt hiệu quả cao trên bộ dữ liệu V-COCO [2], với mong muốn phương thức ứng dụng các đặc trưng mới có thể so sánh được với các SOTA hiện nay.

GIỚI THIỆU

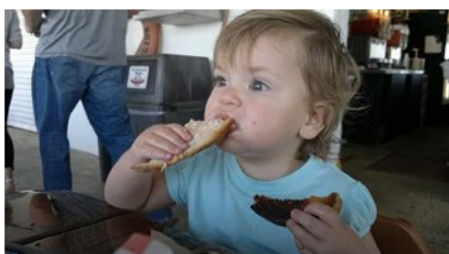
Vấn đề phát hiện tương tác giữa con người và đối tượng là nhằm phát hiện và nhận diện tất cả các tập ba phần tử bao gồm con người, vật thể và tương tác giữa chúng.

Ứng dụng của human object detection rất đa dạng và có thể được sử dụng trong nhiều lĩnh vực khác nhau trong đời sống. Điển hình như, phát hiện hành vi bất thường, hỗ trợ giám sát an ninh và giao thông.

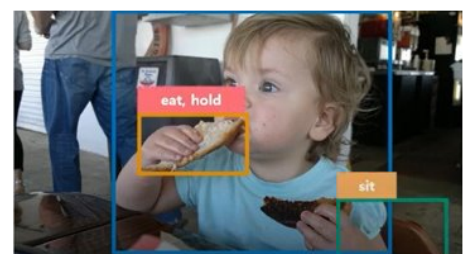
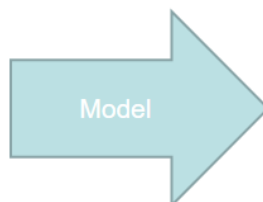
Thông thường, việc phát hiện các thực thể được thực hiện bằng các thuật toán phát hiện đối tượng. Tuy nhiên, việc phát hiện tương tác giữa con người và vật thể là một nhiệm vụ đòi hỏi độ phức tạp cao hơn, yêu cầu mô hình có khả năng nhận diện về cả thông tin ngữ nghĩa trong hình ảnh.

Vì những thách thức cũng như sự phức tạp của bài toán HOI, nhiều nghiên cứu đã được triển khai với mục đích cung cấp một loạt các phương pháp đa dạng nhằm đáp ứng yêu cầu giải quyết vấn đề này. Trong đó, các nghiên cứu đáng chú ý gần đây chủ yếu tập trung vào việc tận dụng kiến trúc transformer [1][3][4], tuy nhiên các nghiên cứu này chỉ tập trung chủ yếu vào sử dụng đặc trưng hình ảnh mà bỏ qua các thông tin có giá trị khác. Để khắc phục nhược điểm này, đề tài tập trung vào việc tìm hiểu tác động của các loại đặc trưng và nghiên cứu một cách tận dụng hiệu quả các loại đặc trưng này nhằm nâng cao độ chính xác của các phương pháp SOTA hiện tại.

Đầu vào của bài toán là một bức ảnh, đầu ra là tập hợp gồm: hộp giới hạn của người, hộp giới hạn của vật thể, nhãn của vật thể, nhãn của tương tác giữa người và vật thể.



Đầu vào



Đầu ra

Hình 1: Đầu vào và đầu ra của bài toán

MỤC TIÊU

- Tìm hiểu và khảo sát các nghiên cứu liên quan trên bài toán phát hiện tương tác người vật.
- Tìm hiểu và phân tích độ hiệu quả của các loại đặc trưng khác nhau khi ứng dụng vào bài toán phát hiện tương tác người vật.
- Đề xuất một phương pháp mới kết hợp thông tin từ các loại đặc trưng khác nhau và có thể áp dụng vào kiến trúc transformer để đạt hiệu quả cao trên bộ dữ liệu V-COCO.

NỘI DUNG VÀ PHƯƠNG PHÁP

- **Nội dung 1: Tìm hiểu về các cách tiếp cận cho bài toán phát hiện tương tác người vật.**
 - **Mục tiêu:** Cung cấp được một cái nhìn tổng quát về tình hình nghiên cứu của bài toán phát hiện tương tác người vật.
 - **Phương pháp thực hiện:** Tìm đọc các bài báo khảo sát (survey paper) trên các tạp chí uy tín, cũng như đọc các bài báo gần đây về phát hiện tương tác trên các hội nghị lớn về xử lý ảnh như: CVPR, ICCV, AAAI.
 - **Sản phẩm:** Tài liệu tổng hợp về tình hình nghiên cứu và tổng quan lý thuyết các phương pháp / cách tiếp cận trên bài toán phát hiện tương tác người vật.
- **Nội dung 2: Tìm hiểu và thực nghiệm các phương pháp phát hiện đối tượng phát hiện tương tác người vật dựa trên transformer trước đó như HOTR, HOITrans, AS-Net. Bên cạnh đó tìm hiểu bộ dữ liệu V-COCO - bộ dữ liệu dùng để thực nghiệm cho bài toán.**
 - **Mục tiêu:**
 - HOTR [1], HOITrans [3], AS-Net [4] đều là các phương pháp điển hình cho việc sử dụng kiến trúc transformer trên bài toán phát hiện tương tác người vật, do đó mục tiêu là nắm ý tưởng chính cách

hoạt động của các phương pháp này.

- Nắm rõ được đặc điểm của bộ dữ liệu V-COCO [1] cùng các thách thức.

- **Phương pháp thực hiện:**

- Tìm đọc các bài báo để tìm hiểu lý thuyết của các phương pháp đề cập, cài đặt lại source code của các phương pháp này.
- Đọc bài báo công bố bộ dữ liệu V-COCO [1].

- **Sản phẩm:**

- Tài liệu tổng hợp lý thuyết của 3 phương pháp đề cập.
- Tài liệu báo cáo tìm hiểu về bộ dữ liệu V-COCO [1].

- **Nội dung 3: Khảo sát, nghiên cứu các đặc trưng trong ảnh có thể sử dụng cho lĩnh vực thị giác.**

- **Mục tiêu:** Quan sát và phân tích được sự hiệu quả khi sử dụng thông tin từ các loại đặc trưng khác nhau.

- **Phương pháp thực hiện:**

- Sử dụng các mô hình pre-trained có sẵn để trích xuất các loại đặc trưng khác nhau.
- Áp dụng các loại đặc trưng này vào các mô hình transformer phát hiện tương tác người vật đã tìm hiểu.
- Quan sát và phân tích tác động của từng loại đặc trưng khi áp dụng vào bài toán phát hiện tương tác người vật.

- **Sản phẩm:**

- Báo cáo và phân tích kết quả huấn luyện khi kết hợp đặc trưng thị giác và các loại đặc trưng khác nhau.
- Bảng kết quả thực nghiệm được đo trên độ đo mAP.

- **Nội dung 4: Đề xuất một cách kết hợp đặc trưng mới: sử dụng kết quả tốt nhất trong các thử nghiệm và đề xuất một cách kết hợp các đặc trưng lại với nhau, gọi là đặc trưng tăng cường.**

- **Mục tiêu:** Đề xuất một thuật toán kết hợp đặc trưng mới mang lại hiệu

quả cao, giúp nâng cao độ chính xác của các phương pháp đã đề cập ở nội dung 2.

- **Phương pháp thực hiện:** Sau khi phân tích hiệu quả của các loại đặc trưng khi áp dụng vào bài toán phát hiện tương tác người và vật, chúng tôi sẽ chạy thử các cách kết hợp chúng lại và quan sát kết quả, sau đó chọn cách kết hợp cho kết quả cao nhất trên bộ dữ liệu V-COCO [1].
- Sản phẩm:
 - Báo cáo và phân tích kết quả huấn luyện khi thử nghiệm các cách kết hợp đặc trưng khác nhau.
 - Bảng kết quả thực nghiệm được đo trên độ đo mAP.

KẾT QUẢ MONG ĐỢI

- Xác định các đặc trưng cho hiệu quả cao khi áp dụng vào mô hình transformer trong bài toán phát hiện tương tác người vật.
- Đề xuất một phương pháp kết hợp đặc trưng mới cho hiệu quả cao khi áp dụng vào mô hình transformer trong bài toán phát hiện tương tác người vật.
- Công bố được bài báo khoa học tại hội nghị quốc tế thuộc danh mục SCOPUS.

TÀI LIỆU THAM KHẢO (Định dạng DBLP)

- [1] Kim, B., Lee, J., Kang, J., Kim, E.-S., & Kim, H. J. (2021). HOTR: End-to-End Human-Object Interaction Detection With Transformers. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2021)* (pp. 74-83). Computer Vision Foundation / IEEE. doi:10.1109/CVPR46437.2021.00014.
- [2] Gkioxari, G., Girshick, R. B., Dollár, P., & He, K. (2018). Detecting and Recognizing Human-Object Interactions. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018* (pp. 8359-8367). Computer Vision Foundation / IEEE Computer Society. doi:10.1109/CVPR.2018.00872
- [3] C. Huang, Q. Wang, and R. Nevatia, "HOITrans: End-to-End Learning of Human-Object Interaction with Transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021*, pp. 15709-15718.
- [4] Chen, M., Liao, Y., Liu, S., Chen, Z., Wang, F., & Qian, C. (2021). Reformulating

HOI Detection As Adaptive Set Prediction. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021 (pp. 9004-9013). Computer Vision Foundation / IEEE. doi:10.1109/CVPR46437.2021.00889.

----- *Trang này cố tình để trống* -----