

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học: CS519 - PHƯƠNG PHÁP LUẬN NCKH

Lớp: CS519.N11

GV: PGS.TS. Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



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

Lê Việt Thịnh - 20520781

Đinh Nhật Minh - 20521597

Dương Thành Bảo Khanh - 20521444

Tóm tắt

- Link Github của nhóm: [Banhkun/CS519.N11 \(github.com\)](https://github.com/Banhkun/CS519.N11)
- Link YouTube video: <https://www.youtube.com/watch?v=Gv8zh8FDGH4>



LÊ VIỆT THỊNH



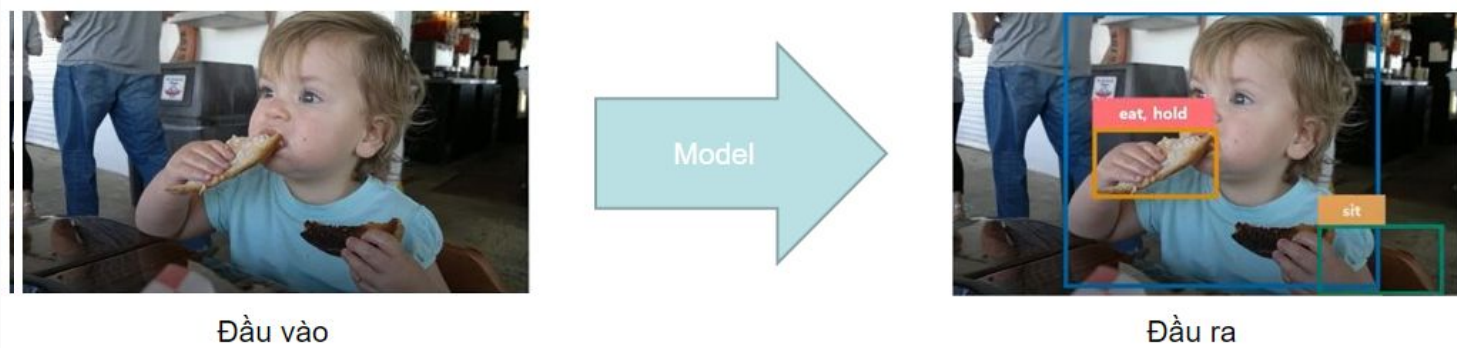
ĐINH NHẬT MINH



DƯƠNG THÀNH BẢO KHANH

Giới thiệu

- Bài toán: phát hiện tương tác người và vật (HOI)



- Phát hiện đối tượng + thông tin ngữ nghĩa trong ảnh
- Nghiên cứu gần đây: Transformer
- Đề tài: nghiên cứu cách tận dụng các loại đặc trưng → nâng cao độ chính xác của các phương pháp SOTA

Mục tiêu

- Tìm hiểu và khảo sát các nghiên cứu liên quan trên bài toán
- Tìm hiểu và phân tích độ hiệu quả của các loại đặc trưng khác nhau khi ứng dụng vào bài toán
- 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.
 - Cách thực hiện: tìm đọc các bài báo khảo sát trên các tạp chí uy tín, các bài báo gần đây trên các hội nghị lớn.
- 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; tìm hiểu bộ dữ liệu V-COCO.
 - Cách thực hiện: Đọc các bài báo để tìm hiểu lý thuyết của các phương pháp, cài đặt lại source code của các phương pháp; đọc bài báo công bố bộ dữ liệu V-COCO.

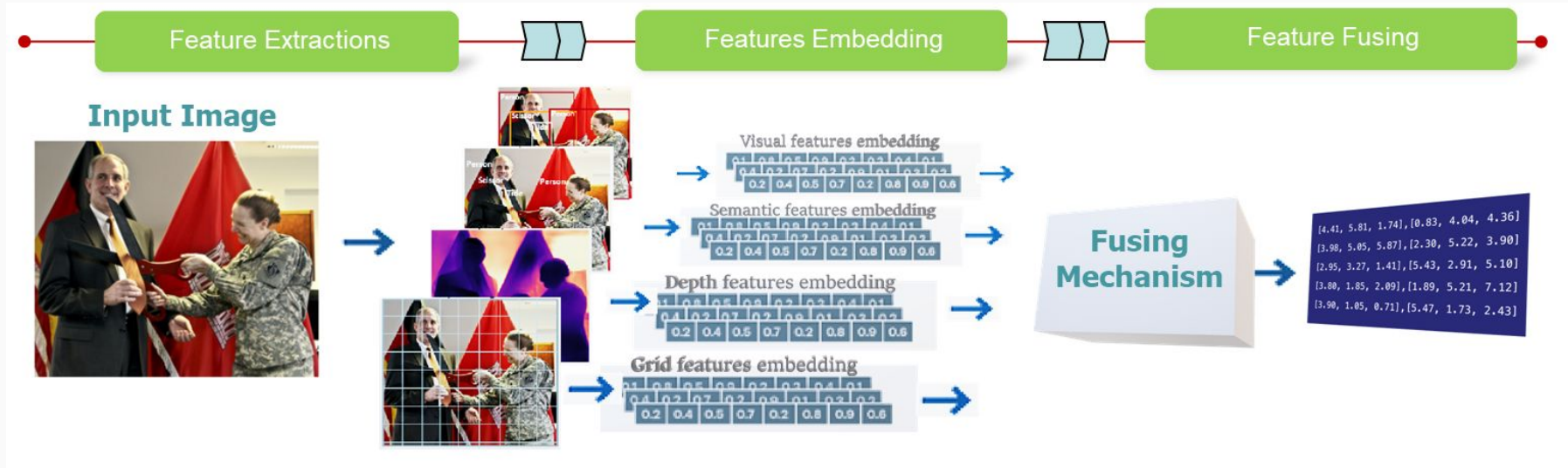
Nội dung và Phương pháp

- Nội dung 3: Khảo sát, nghiên cứu các đặc trưng trong ảnh có thể sử dụng cho lĩnh vực thị giác.
 - 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 đã tìm hiểu.
 - Quan sát và phân tích tác động của từng loại đặc trưng khi áp dụng vào bài toán.

Nội dung và Phương pháp

- Nội dung 4: Đề xuất một cách kết hợp đặc trưng mới: sử dụng kết quả tốt nhất trong các thử nghiệm; đề xuất một cách kết hợp các đặc trưng lại với nhau: đặc trưng tăng cường.
 - Phương pháp thực hiện: Phân tích hiệu quả của các loại đặc trưng khi áp dụng vào bài toán, 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.

Nội dung và Phương pháp



Kết quả dự kiến

- 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

- [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.
-