THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút): *https://www.youtube.com/watch?v=VZz0YaoAU7Y*
- Link slides (dạng .pdf đặt trên Github của nhóm): *https://github.com/lnmduc2/CS519.O11/blob/main/slides.pdf*

• Họ và Tên: · Lê Ngô Minh Đức • MSSV: 21520195 	• Lớp: CS519.O11 • Tự đánh giá (điểm tổng kết môn): 9/10 • Số buổi vắng: 0 • Số câu hỏi QT cá nhân: ?? • Số câu hỏi QT của cả nhóm: ?? • Link Github: https://github.com/lnmduc2/CS519.O11 • Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: ○ Lên ý tưởng cho đồ án ○ Viết đề cương, chỉnh sửa Slide và Poster ○ Làm video YouTube

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

PHƯƠNG PHÁP THỊ GIÁC MÁY TÍNH ĐỔI MỚI PHỤC VỤ HỆ THỐNG THANH TOÁN TỰ ĐỘNG

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

INNOVATIVE VISION-BASED METHOD FOR AUTOMATED CHECKOUT SYSTEMS

TÓM TẮT

Trong lĩnh vực bán lẻ, sự phát triển nhanh chóng và đáng kể đã được thúc đẩy bởi sự kết hợp giữa công nghệ thông tin và tương tác khách hàng, với mô hình Amazon Go là ví dụ nổi bật. Tuy nhiên, việc giữ bí mật thông tin công nghệ, nhất là từ các công ty lớn lo ngại về vấn đề bản quyền, đã tạo ra thách thức trong việc nghiên cứu chi tiết và phát triển ứng dụng của xử lý ảnh và học sâu trong các hệ thống thanh toán tự động. Hơn nữa, các nghiên cứu hiện hành chủ yếu tập trung vào việc cải thiện các khung làm việc (framework) sẵn có, nhưng lại thiếu tính phổ quát khi áp dụng vào môi trường bán lẻ thực tế. Do đó, việc nghiên cứu để phát triển một giải pháp tổng quát hơn trở nên cấp thiết. Trong đề tài này, nhóm không chỉ cố gắng đề xuất một hệ thống thị giác máy tính mới trong lĩnh vực bán lẻ tự động, mà còn hướng tới việc tạo ra giải pháp thực tế hơn, có khả năng ứng dụng rộng rãi trong ngành.

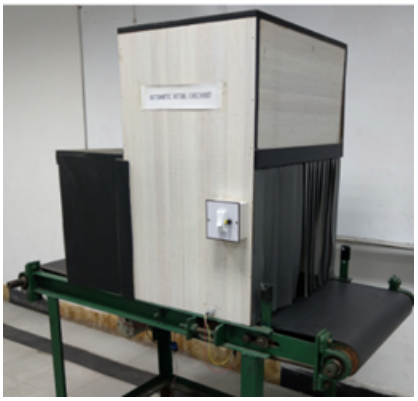
GIỚI THIỆU

Trong thế giới ngày càng tiên tiến về công nghệ, ngành bán lẻ đã chứng kiến những bước tiến vượt bậc nhờ vào sự hợp nhất của công nghệ thông tin và tương tác khách hàng. Amazon, tiên phong trong lĩnh vực này, đã giới thiệu cửa hàng Amazon Go vào năm 2018 tại Seattle, loại bỏ hoàn toàn quá trình xếp hàng và thanh toán truyền thống nhờ vào hệ thống AI và camera, đánh dấu bước tiến lớn trong ứng dụng công nghệ thị giác máy tính và AI trong bán lẻ. Điểm nổi bật của mô hình cửa hàng Amazon Go là công nghệ "Just Walk Out", cho phép khách hàng chỉ cần lấy sản phẩm họ muốn và rời cửa hàng mà không cần qua quá trình thanh toán truyền thống, với hệ thống camera và cảm biến tự động tính tiền và ghi nhận vào tài khoản của họ.

Hình 1: Công nghệ "Just Walk Out" của Amazon, cho phép khách hàng chỉ cần lấy sản phẩm và rời đi mà không cần xếp hàng hay quét mã sản phẩm.

Tuy nhiên, ngành công nghiệp này đối mặt với thách thức lớn do sự thiếu hụt trong nghiên cứu chi tiết và công bố về ứng dụng của xử lý ảnh và học sâu trong các hệ thống thanh toán tự động. Lí do chính cho sự thiếu hụt này xuất phát từ việc các công ty lớn như Amazon giữ kín thông tin về công nghệ của họ nhằm bảo vệ bản quyền và lợi ích kinh doanh, khiến cho việc hiểu biết sâu rộng về cách thức hoạt động và áp dụng công nghệ thị giác máy tính trong lĩnh vực bán lẻ trở nên khó khăn.

Với nỗ lực giải quyết vấn đề trên, các nghiên cứu gần đây phần lớn tập trung sử dụng các công nghệ như mạng thần kinh tích chập hay LSTM để nâng cao hiệu quả nhận dạng hành động của khách hàng. Tuy nhiên, các nghiên cứu này chỉ tập trung cải tiến các framework trước đó, và thường được thực hiện trong môi trường bán lẻ đã được điều chỉnh, chưa phản ánh đúng mức độ phổ quát cần thiết của một cửa hàng thực tế. Bởi vì, ngoài việc quan sát hành động khách hàng, việc xác định sản phẩm lấy đi hay trả lại hoặc chọn lựa góc quay tối ưu cũng là những yếu tố quan trọng cần được xem xét. Ví dụ, các bài báo [1] và [2], mặc dù tập trung vào nhận dạng tư thế thông qua việc phân loại hành động bằng mạng LSTM và có quan tâm đến vấn đề quyền riêng tư của khách hàng, vẫn thiếu sự xem xét sâu rộng về các yếu tố thực tiễn cần thiết cho một hệ thống bán lẻ hoàn chỉnh. Hoặc gần đây hơn, bài báo [3] vẫn còn hạn chế lớn là giả định về thiết kế hệ thống không phản ánh chính xác điều kiện thực tế trong môi trường bán lẻ. Cụ thể, hệ thống được thiết kế để xử lý từng sản phẩm riêng lẻ và yêu cầu kích thước sản phẩm phù hợp với trường nhìn của webcam, điều này hạn chế hiệu quả của nó trong các cửa hàng có đa dạng sản phẩm và kích thước.



Hình 2: Thiết kế mẫu của hệ thống đề xuất bởi bài báo [3] ("ARC: A Vision-based Automatic Retail Checkout System")

Chính vì sự thiếu hụt nghiên cứu chi tiết và ứng dụng thực tế trong công nghệ thị giác máy tính bán lẻ, cùng với công nghệ "Just Walk Out" tiên tiến của Amazon, đã đặt ra **câu hỏi**: “Làm thế nào để tìm ra giải pháp tổng quát hơn, phù hợp với đa dạng các cửa hàng bán lẻ, không chỉ giới hạn ở việc cải tiến hiệu suất các phương pháp hiện có?”. Vậy nên, đề tài của nhóm sẽ tập trung vào việc xây dựng và tích hợp một phương pháp thông minh có khả năng phát hiện sản phẩm nào được lấy ra khỏi kệ hàng hoặc được đặt trở lại kệ hàng, và chuyển giao mặt hàng đó sang khách hàng để phục vụ thanh toán tự động. Điều này mở ra hướng nghiên cứu và phát triển mới trong lĩnh vực bán lẻ, tạo ra các lựa chọn thực tiễn hơn cho ngành công nghiệp nói trên. Vậy bài toán có:

Input: Hình ảnh từ Camera ghi lại hình ảnh sản phẩm và khách hàng trong quá trình mua sắm, và thông tin chi tiết của từng sản phẩm như mã sản phẩm, giá cả, và các đặc tính khác.

Output: Hóa đơn điện tử với tất cả thông tin chi tiết về giao dịch, gửi đến khách hàng qua email hoặc ứng dụng di động.

MỤC TIÊU

- Tìm hiểu, huấn luyện mô hình YOLOv8 để nhận diện chính xác danh tính khách hàng cùng sản phẩm trên kệ hàng, và sử dụng mô hình nhận dạng tư thế con người (như OpenPifPaf) để phân tích hành vi khách hàng từ góc nhìn camera giám sát trên cao.

- Xây dựng quy trình thị giác máy tính để theo dõi và phát hiện sự thay đổi của sản phẩm trên kệ hàng, bao gồm việc sản phẩm nào biến mất hoặc được thêm lại bằng các thuật toán xử lí ảnh đơn giản.

- Xây dựng quy trình chuyển giao thông tin sản phẩm từ việc phát hiện sản phẩm được lấy ra hoặc đặt trở lại kệ hàng sang khách hàng tương ứng, dựa trên kết quả từ mô hình thị giác máy tính và nhận dạng tư thế.

NỘI DUNG VÀ PHƯƠNG PHÁP

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

Nội dung 1. Tìm hiểu, huấn luyện mô hình YOLOv8 để nhận diện chính xác danh tính khách hàng cùng sản phẩm trên kệ hàng, và sử dụng mô hình nhận dạng tư thế con người (như OpenPifPaf) để phân tích hành vi khách hàng trong môi trường mua sắm từ góc nhìn camera giám sát trên cao.

Phương pháp thực hiện:

+ Huấn luyện mô hình YOLOv8 để nhận diện chính xác danh tính khách hàng cùng sản phẩm trên kệ hàng từ góc nhìn camera giám sát từ trên cao:

· Xây dựng một bộ dữ liệu thủ công với góc nhìn camera giám sát từ trên cao, vì các bộ dữ liệu trên mạng hầu như đều có kích thước sản phẩm trong 1 bức ảnh là khá lớn do góc nhìn camera gần, không phù hợp với bài toán nhận dạng sản phẩm trong môi trường cửa hàng bán lẻ, siêu thị. Tập dữ liệu này sẽ được gán nhãn bằng tool LabelImg [4].

· Tìm hiểu về mô hình YOLOv8 [5] và độ đo mAP sử dụng đánh giá độ tốt của mô hình với bộ dữ liệu trên, vì nó có khả năng real-time và độ chính xác khá tốt trong thực tế.

· Tìm hiểu và thí nghiệm những phiên bản nào của YOLOv8 (nano, small, medium, large and extra large) có khả năng phát hiện các sản phẩm với độ chính xác cao và cân đối với tốc độ xử lý để có thể giám sát thời gian thực.

· Tìm hiểu mô hình ByteTrack [6] vì thuật toán này thuộc dạng Tracking-by-detection, nó cần sử dụng kết quả của bài toán object detection chính là bounding box bao quanh vật thể thu được từ YOLOv8. Sau đó, tìm cách liên kết các bounding box thu được ở mỗi frame và gán ID cho từng đối tượng.

+ Sử dụng mô hình nhận dạng tư thế con người (như OpenPifPaf) để phân tích hành vi khách hàng trong môi trường mua sắm từ góc nhìn camera giám sát từ trên cao:

- Tìm hiểu về model OpenPifPaf [7] (một công cụ ước tính tư thế nhiều người 2D, nhưng được thiết kế đặc biệt dành cho các chuỗi video thay vì hình ảnh tĩnh) vì thuật toán này được thiết kế để xử lý tốt với những tư thế con người phức tạp và trong các tình huống có sự che khuất, phù hợp trong môi trường mua sắm.

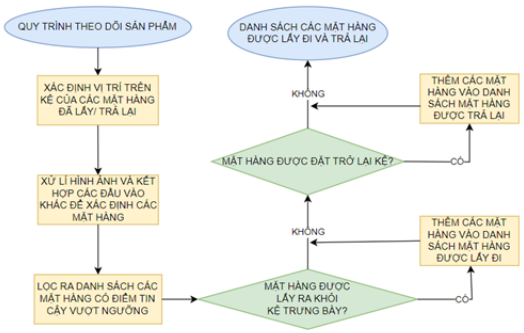
- Tìm hiểu về tập dữ liệu MPII Human Pose (do đây là một trong những bộ dữ liệu chuẩn được sử dụng rộng rãi trong nghiên cứu và phát triển về nhận dạng tư thế), huấn luyện với mô hình OpenPifPaf và đánh giá bằng độ đo AP (Average Precision).

Nội dung 2. Xây dựng quy trình thị giác máy tính để theo dõi và phát hiện sự thay đổi của sản phẩm trên kệ hàng, bao gồm việc sản phẩm nào biến mất hoặc được thêm lại bằng các thuật toán xử lí ảnh đơn giản.

Phương pháp thực hiện:

- Tìm hiểu các thuật toán xử lí ảnh đơn giản để phát hiện sự thay đổi trên kệ hàng trong một chuỗi các hình ảnh như Optical Flow, Consecutive Frames Subtraction hoặc Background Subtraction (**Lí do:** việc sử dụng những thuật toán này sẽ giảm bớt tài nguyên tính toán cần thiết so với các thuật toán tracking phức tạp, khi mà phần lớn sản phẩm không di chuyển thường xuyên. Các phương pháp này cung cấp một cách tiếp cận hiệu quả và tinh gọn hơn để theo dõi sự thay đổi trên kệ hàng, giúp quản lý hàng tồn kho một cách thông minh và tiết kiệm tài nguyên máy chủ).

- Xây dựng quy trình theo dõi và phát hiện sự thay đổi của sản phẩm trên kệ hàng theo sơ đồ giản lược:

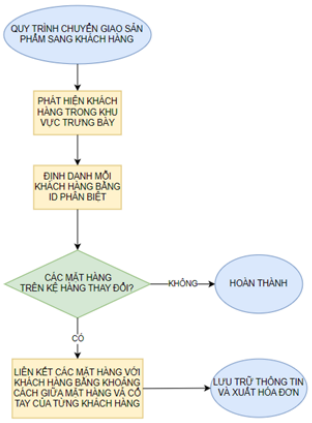


Hình 3: Sơ đồ giản lược cho quy trình theo dõi sản phẩm

Nội dung 3. Xây dựng quy trình chuyển giao thông tin sản phẩm từ việc phát hiện sản phẩm được lấy ra hoặc đặt trở lại kệ hàng sang khách hàng tương ứng, dựa trên kết quả từ mô hình thị giác máy tính và nhận dạng tư thế.

Phương pháp thực hiện:

Xây dựng quy trình chuyển giao thông tin sản phẩm từ việc phát hiện sản phẩm được lấy ra hoặc đặt trở lại kệ hàng sang khách hàng tương ứng theo sơ đồ giản lược:



Hình 4: Sơ đồ giản lược cho quy trình chuyển giao sản phẩm sang khách hàng

KẾT QUẢ MONG ĐỢI

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

- Tập dữ liệu được quay từ camera góc cao, gồm ít nhất 2500 ảnh, mỗi ảnh có thể chứa khoảng 1 đến 2 người và có khoảng 5000 tư thế.

- Mô hình nhận diện danh tính khách hàng cùng sản phẩm trên kệ hàng và mô hình nhận dạng tư thế con người đạt kết quả cao và có khả năng real-time.

- Xây dựng được quy trình phát hiện sự thay đổi của sản phẩm trên kệ hàng và quy trình chuyển giao thông tin sản phẩm sang khách hàng theo đúng sơ đồ giản lược.

TÀI LIỆU THAM KHẢO

[1] Varadarajan, Srenivas, and Shahrokh Shahidzadeh. "A Real-Time System for Shoppers' Action Recognition." Electronic Imaging 2016.3 (2016): 1-6.

[2] Moghaddam, Mohammad Mahdi Kazemi, Ehsan Abbasnejad, and Javen Shi. "Follow the attention: Combining partial pose and object motion for fine-grained action detection." arXiv preprint arXiv:1905.04430 (2019).

[3] Bukhari, Syed Talha, et al. "Arc: A vision-based automatic retail checkout system." arXiv preprint arXiv:2104.02832 (2021).

[4] https://github.com/HumanSignal/labelImg

[5] https://github.com/ultralytics/ultralytics

[6] Zhang, Yifu, et al. "Bytetrack: Multi-object tracking by associating every detection box." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

[7] Kreiss, Sven, Lorenzo Bertoni, and Alexandre Alahi. "Openpifpaf: Composite fields for semantic keypoint detection and spatio-temporal association." IEEE Transactions on Intelligent Transportation Systems 23.8 (2021): 13498-13511.