

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ

**BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

**XÂY DỰNG ỨNG DỤNG DI ĐỘNG
GIỚI THIỆU ÂM THỰC MIỀN TÂY NAM BỘ**

THS2022-17

Cần Thơ, 11/2022

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ

**BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

**XÂY DỰNG ỨNG DỤNG DI ĐỘNG
GIỚI THIỆU ÂM THỰC MIỀN TÂY NAM BỘ**

THS2022-17

Sinh viên thực hiện: Nguyễn Thị Mỹ Khánh Nam, Nữ: Nữ

Dân tộc: Kinh

Lớp, khoa: Lớp DI191V7F2, Khoa CNTT Năm thứ: 4 Số năm đào tạo: 4,5

Ngành học: Công Nghệ Thông Tin

Người hướng dẫn: TS. Lâm Nhựt Khang

Cần Thơ, 11/2022

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

Giảng viên hướng dẫn: TS. Lâm Nhựt Khang

Chủ nhiệm đề tài: Nguyễn Thị Mỹ Khanh

Thành viên thực hiện:

Nguyễn Duy Khang

Nguyễn Hiếu Nghĩa

Lê Hải Yên

Tống Phước Lộc

LỜI CẢM ƠN

Để hoàn thành được bài báo cáo nghiên cứu khoa học này, nhóm nghiên cứu chân thành cảm ơn sâu sắc đến cô Lâm Nhựt Khang - người đã tận tình hướng dẫn giúp đỡ chúng tôi. Trong quá trình nghiên cứu, nhờ sự chỉ dạy và hướng dẫn mà nhóm nghiên cứu nói chung cũng như bài nghiên cứu khoa học nói riêng này được hoàn thành một cách tốt nhất.

Nhóm nghiên cứu xin gửi lời cảm ơn đến trường Đại học Cần Thơ, các phòng ban hỗ trợ các thủ tục và đã giúp sức về mặt chi phí để nhóm có thể hoàn thành một cách tốt nhất.

Nhóm nghiên cứu cũng xin gửi lời cảm ơn chân quý đến các Thầy Cô giảng viên Đại học Cần Thơ, đặc biệt là các Thầy cô ở Trường Công nghệ Thông tin và Truyền thông, những người đã mang đến chúng em những kiến thức trong thời gian theo học tại khoa.

Nhóm nghiên cứu cũng gửi một chút lời cảm ơn đến bạn bè và gia đình đã luôn ủng hộ, động viên và tạo điều kiện tốt nhất để nhóm có thể hoàn thành công trình nghiên cứu một cách toàn vẹn.

Tuy cố gắng trong quá trình thực hiện nhưng bên cạnh đó cũng không thể tránh được những sai sót. Nhóm cũng mong nhận được sự góp ý kiến đến từ quý Thầy Cô và các bạn để có được một bài báo cáo hoàn thiện hơn.

Cần Thơ, ngày tháng năm 2022
Chủ nhiệm đề tài

Nguyễn Thị Mỹ Khánh

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

STT	Họ và tên	MSSV, Lớp, Khoa, Khóa	Nội dung nghiên cứu cụ thể được giao
1	Nguyễn Thị Mỹ Khánh	B1910657 Lớp: Công nghệ thông tin – Chất lượng cao 2 Khóa: 45	- Nghiên cứu lý thuyết về các mô hình nhận dạng đối tượng và các nghiên liên quan đến ứng dụng di động - Thu nhập và xây dựng tập dữ liệu mô hình - Viết báo cáo tổng kết đề tài
2	Nguyễn Duy Khang	B1910654 Lớp: Công nghệ thông tin – Chất lượng cao 2 Khóa: 45	- Nghiên cứu lý thuyết về các mô hình nhận dạng đối tượng và các nghiên liên quan đến ứng dụng di động - Thu nhập và xây dựng tập dữ liệu mô hình - Viết báo cáo tổng kết đề tài
3	Nguyễn Hiếu Nghĩa	B1910672 Lớp: Công nghệ thông tin – Chất lượng cao 2 Khóa: 45	- Thu nhập và xây dựng tập dữ liệu mô hình - Lập trình, xây dựng, thiết kế ứng dụng di động - Viết báo cáo tổng kết đề tài
4	Lê Hải Yến	B1910731 Lớp: Công nghệ thông tin – Chất lượng cao 2 Khóa: 45	- Thu nhập và xây dựng tập dữ liệu mô hình - Lập trình, xây dựng, thiết kế ứng dụng di động - Viết báo cáo tổng kết đề tài
5	Tống Phước Lộc	B1910664 Lớp: Công nghệ thông tin – Chất lượng cao 2 Khóa: 45	- Thu nhập và xây dựng tập dữ liệu mô hình - Lập trình, xây dựng, thiết kế ứng dụng di động

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

		- Viết báo cáo tổng kết đề tài
--	--	-----------------------------------

MỤC LỤC

DANH MỤC BẢNG BIỂU	3
DANH MỤC HÌNH ẢNH	4
DANH MỤC TỪ VIẾT TẮT	6
LỜI MỞ ĐẦU	12
CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN	13
1.1. Cơ sở khoa học và thực tiễn của đề tài	13
1.2. Những nghiên cứu liên quan	13
1.3. Mục tiêu	15
1.4. Đối tượng và phạm vi nghiên cứu	15
1.5. Phương pháp nghiên cứu	16
1.6. Bố cục	16
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT	17
2.1. Mô hình YOLO	17
2.1.1. Tổng quan YOLO	17
2.1.2. Mô hình YOLOv5	18
2.1.3. Mô hình YOLOv7	19
2.1.4. Output của YOLO	22
2.2. Mô hình Vision Transformer	22
2.3. Mô hình MobileNet-v3	27
2.4. Phương pháp đánh giá precision, recall, AP và mAP	29
CHƯƠNG 3. PHƯƠNG PHÁP THỰC HIỆN	32
3.1. Mô hình YOLO	32
3.2. Mô hình Vision Transformer	33

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ

3.3. Mô hình MobileNet-v3	34
CHƯƠNG 4. XÂY DỰNG TẬP DỮ LIỆU	35
4.1. Thu thập dữ liệu	35
4.2. Tiền xử lý dữ liệu	36
4.2.1. Tách tập dữ liệu	36
4.2.2. Tăng cường dữ liệu	37
4.2.3. Chuyển đổi dữ liệu sang định dạng YOLO	41
CHƯƠNG 5. THỰC NGHIỆM VÀ ỨNG DỤNG	42
5.1. Thiết lập dữ liệu và tổ chức thư mục	42
5.2. Huấn luyện mô hình YOLOv5	42
5.3. Triển khai mô hình bằng TensorFlow Lite	48
CHƯƠNG 6. KẾT LUẬN	54
6.1. Kết quả đạt được	54
6.2. Hướng phát triển	54
TÀI LIỆU THAM KHẢO	55

DANH MỤC BẢNG BIỂU

Bảng 1 . Số lượng hình ảnh trong dataset sau khi sử dụng các kỹ thuật tăng cường dữ liệu	41
Bảng 2 . Các Hyperparameters mặc định và được sửa đổi để huấn luyện mô hình YOLOv5	43
Bảng 3 . Kết quả huấn luyện mô hình YOLOv5	44
Bảng 4 . mAP của các mô hình nhận dạng món ăn MobileNet-V3(MoNV3), YOLOv5, YOLOv7, và Vision Transformer(ViT)	48

DANH MỤC HÌNH ẢNH

<i>Hình 1 . Kiến trúc mô hình YOLO [19]</i>	17
<i>Hình 2 . YOLO timeline</i>	18
<i>Hình 3 . Các phiên bản của YOLOv5</i>	18
<i>Hình 4 . Backbone của YOLOv7 [24]</i>	20
<i>Hình 5 . Neck của YOLOv7[24]</i>	21
<i>Hình 6 . Kiến trúc mô hình Vision Transformer [25]</i>	22
<i>Hình 7 . Cùng một ảnh đầu vào và cách chia patch nhưng thứ tự các patch khác nhau</i>	23
<i>Hình 8 . Khởi tạo các vector q, k, v</i>	25
<i>Hình 9 . Nhân lần lượt các vector q với vector k</i>	25
<i>Hình 10 . Chia score cho căn bậc 2 kích thước vector k rồi đưa qua hàm softmax</i>	26
<i>Hình 11 . Nhân giá trị softmax với vector v sau đó cộng các vector kết quả lại</i>	27
<i>Hình 12 . Mô hình SE-ResNet Module[28]</i>	28
<i>Hình 13 . MobileNet-V3 block [26]</i>	29
<i>Hình 14 . Công thức tính IoU</i>	30
<i>Hình 15 . Giá trị IoU cho bounding box</i>	31
<i>Hình 16 . Mô hình YOLO5 cho bài toán nhận diện món ăn Tây Nam Bộ</i>	32
<i>Hình 17 . Mô hình Vision Transformer cho bài toán nhận diện món ăn Tây Nam Bộ.</i>	33
<i>Hình 18 . Mô hình MobileNet-V3 cho bài toán nhận diện món ăn Tây Nam Bộ</i>	34
<i>Hình 19 . Ví dụ một số hình ảnh từ tập dữ liệu</i>	36
<i>Hình 20 . Gán nhãn thủ công với Roboflow</i>	36
<i>Hình 21 . Dữ liệu sau khi được phân chia</i>	37
<i>Hình 22 . Minh họa kỹ thuật Flip</i>	38
<i>Hình 23 . Minh họa kỹ thuật 90° Rotate</i>	38

<i>Hình 24 . Minh họa kỹ thuật Blur</i>	39
<i>Hình 25 . Cấu trúc thư mục dữ liệu.....</i>	42
<i>Hình 26 . Lables correlogram.....</i>	45
<i>Hình 27 . Lables</i>	45
<i>Hình 28 . Kết quả train</i>	46
<i>Hình 29 . Ma trận nhầm lẫn.....</i>	46
<i>Hình 30 . Giao diện trang chủ của ứng dụng với tiếng Việt.....</i>	49
<i>Hình 31 . Giao diện trang chủ của ứng dụng với tiếng Anh</i>	49
<i>Hình 32 . Giao diện khi chọn chức năng “Nhận dạng trực tiếp”</i>	49
<i>Hình 33 . Giao diện khi chọn chức năng “Live Detection”</i>	49
<i>Hình 34 . Giao diện khi chọn chức năng “Nhận dạng qua ảnh”</i>	50
<i>Hình 35 . Giao diện khi chọn chức năng “Gallery Detection”</i>	50
<i>Hình 36 . Giao diện ứng dụng khi chọn chức năng “Tải ảnh”</i>	50
<i>Hình 37 . Giao diện ứng dụng khi chọn chức năng “Upload Image”</i>	51
<i>Hình 38 . Giao diện ứng dụng khi chọn chức năng “Camera” với tiếng Việt.....</i>	51
<i>Hình 39 . Giao diện ứng dụng khi chọn chức năng “Camera” với tiếng Anh.....</i>	52
<i>Hình 40 . Giao diện ứng dụng khi chọn chức năng “Thông tin món ăn”</i>	52
<i>Hình 41 . Giao diện ứng dụng khi chọn chức năng “Detail Food”</i>	53
<i>Hình 42 . Giao diện ứng dụng chức năng tìm kiếm với tiếng Việt.....</i>	53
<i>Hình 43 . Giao diện ứng dụng chức năng tìm kiếm với tiếng Anh</i>	53

DANH MỤC TỪ VIẾT TẮT

No.	Abbreviation	Origin word
1	CNN	Convolutional Neural Network
2	DCNN	Deep Convolutional Neural Network
3	PRENet	Progressive Region Enhancement Network
4	SPPNet	Spatial Pyramid Pooling Networks
5	YOLO	You Only Look Once
6	ViT	Vision Transformer
7	CSP	Cross Stage Partial Networks
8	ELAN	Efficient Layer Aggregation Network
9	NLP	Natural Language
10	MSA	Multi-Head Self-Attention
11	MLP	Multi-Layer Perceptron
12	LN	Layernorm
13	LR-ASPP	Lite Reduced Atrous Spatial Pyramid Pooling
14	SE	Squeeze and Excitation
15	IoU	Intersection over Union
16	AP	Average Precision
17	mAP	Mean Average Precision
18	MoNV3	MobileNet-V3

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC CẦN THƠ

THÔNG TIN KẾT QUẢ NGHIÊN CỨU CỦA ĐỀ TÀI

1. Thông tin chung:

- Tên đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ
- Sinh viên thực hiện: Nguyễn Thị Mỹ Khánh
- Lớp: DI19V7F2 Trường: CNTT&TT Năm thứ: 4 Số năm đào tạo: 4,5
- Người hướng dẫn: TS. Lâm Nhựt Khang

2. Mục tiêu đề tài:

Mục tiêu đề tài là xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ, và cho phép nhận diện một số món ăn đặc trưng của miền Tây Nam Bộ.

3. Tính mới và sáng tạo:

Đề tài nghiên cứu các mô hình thị giác máy tính đương đại và ứng dụng được vào thực tiễn, góp phần xây dựng nền du lịch địa phương cũng như du lịch nước nhà.

4. Kết quả nghiên cứu:

Nhóm đã thu thập được tập dữ liệu và xây dựng thành công ứng dụng di động cho phép nhận diện một số món ăn đặc trưng của miền Tây Nam Bộ.

5. Đóng góp về mặt kinh tế - xã hội, giáo dục và đào tạo, an ninh, quốc phòng và khả năng áp dụng của đề tài:

Đề tài nghiên cứu khoa học này góp phần quan trọng trong việc nhận dạng món ăn Tây Nam Bộ, hỗ trợ quảng bá du lịch cho các tỉnh Đồng bằng Sông Cửu Long.

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

6. Công bố khoa học của sinh viên từ kết quả nghiên cứu của đề tài, hoặc nhận xét, đánh giá của cơ sở đã áp dụng các kết quả nghiên cứu:

Số thứ tự	Tên bài báo	Tác giả/nhóm tác giả	Tên tạp chí	Số tạp chí	Năm xuất bản
1.	SWVie-Food: A Dataset for Recognizing Foods in Southwest Vietnam Based on Deep Learning	Khang Nhut Lam, My-Khanh Thi Nguyen, Khang Duy Nguyen, Nghia Hieu Nguyen, Kim-Yen Thi Nguyen, Andrew Ware	The 16 th International Conference on Computing and Communication Technologies (RIVF2022)		2022

Ngày tháng năm 2022

Sinh viên chịu trách nhiệm chính

thực hiện đề tài

Nhận xét của người hướng dẫn về những đóng góp khoa học của sinh viên thực hiện đề tài (phản này do người hướng dẫn ghi):

Nhóm nghiên cứu rất chịu khó trong nghiên cứu, chăm chỉ, siêng năng và chủ động trong mọi hoạt động. Đề tài nghiên cứu các mô hình thị giác máy tính đương đại và ứng dụng được vào thực tiễn. Đề tài nghiên cứu khoa học này góp phần quan trọng trong việc nhận dạng món ăn Tây Nam Bộ, hỗ trợ quảng bá du lịch cho các tỉnh Đồng bằng Sông Cửu Long. Và quan trọng hơn, đề tài đã xây dựng được tập dữ liệu các món ăn đặc trưng Tây

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

Nam Bộ, đang trong giai đoạn mở rộng và hoàn thiện để công bố rộng rãi, sử dụng cho các hoạt động nghiên cứu trong tương lai.

Ngày tháng năm 2022

Xác nhận của Trường Đại học Cần Thơ

Người hướng dẫn

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC CẦN THƠ

THÔNG TIN VỀ SINH VIÊN

CHỊU TRÁCH NHIỆM CHÍNH THỰC HIỆN ĐỀ TÀI

I. SƠ LƯỢC VỀ SINH VIÊN:

Họ và tên: Nguyễn Thị Mỹ Khanh

Sinh ngày: 04 tháng: 05 năm: 2001

Nơi sinh: An Giang

Lớp: DI19V7F2 Khóa: 45

Trường: Công nghệ thông tin và Truyền thông

Địa chỉ liên hệ: Cách mạng tháng 8, quận Bình Thủy, tp. Cần Thơ

Điện thoại: 0976038762 Email: khanhb1910657@student.ctu.edu.vn

Ảnh 4x6

II. QUÁ TRÌNH HỌC TẬP (kê khai thành tích của sinh viên từ năm thứ 1 đến năm đang học):

* Năm thứ 1:

Ngành học: Công nghệ thông tin-CLC Trường: Công nghệ thông tin & Truyền thông

Kết quả xếp loại học tập: Giỏi

Sơ lược thành tích:

* Năm thứ 2:

Ngành học: Công nghệ thông tin-CLC Trường: Công nghệ thông tin & Truyền thông

Kết quả xếp loại học tập: Khá

Sơ lược thành tích:

* Năm thứ 3:

Ngành học: Công nghệ thông tin-CLC Trường: Công nghệ thông tin & Truyền thông

Kết quả xếp loại học tập: Giỏi

Sơ lược thành tích:

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

Ngày tháng năm 2022

Xác nhận của Trường Đại học Cần Thơ

**Sinh viên chịu trách nhiệm chính
thực hiện đề tài**

LỜI MỞ ĐẦU

Trong nhiều năm trở lại đây cùng với sự bùng nổ của Internet, cuộc cách mạng công nghiệp 4.0 đang lan rộng trên hầu hết mọi lĩnh vực đời sống, công nghệ thông tin đã và đang dần thay thế để hỗ trợ con người trong một số công việc giúp mang lại hiệu quả cao hơn, tránh sai sót đáng có. Nhiều nền tảng công nghệ mới được tạo ra và ứng dụng vào thực tế như trí tuệ nhân tạo, xử lý dữ liệu lớn, máy học,... Mạng nơ ron học sâu là lĩnh vực nghiên cứu các thuật toán, chương trình máy tính để máy tính học tập và đưa ra những dự đoán như con người.

Tại Việt Nam, việc áp dụng các kỹ thuật phát hiện đối tượng đang trở nên đa dạng như nhận dạng vật thể giúp thanh toán hàng hóa nhanh, áp dụng trong hỗ trợ nông trại, ... đây đều là các hướng nghiên cứu mới góp phần trong việc đem lại hiệu quả cao, tối ưu các khâu trong các lĩnh vực. Ngày nay, trí tuệ nhân tạo được ứng dụng vào nhiều lĩnh vực khác nhau như khoa học, kỹ thuật, đời sống cũng như các ứng dụng về phân loại và phát hiện đối tượng. Nghiên cứu này trình bày kết quả triển khai mô hình thị giác máy tính trong việc nhận dạng các món ăn đặc trưng của miền tây Nam Bộ. Kết quả cho thấy, YOLOv5 là một giải pháp đơn giản, dễ dàng chuyển đổi và thực thi trên các thiết bị có cấu hình thấp. Điều này có thể mở ra hướng tiếp cận mới trong việc nhận dạng các đối tượng Realtime trên thiết bị Smartphone.

CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN

Chương 1 sẽ trình bày tổng quan đề tài, các nghiên cứu có liên quan, mục tiêu, đối tượng, phạm vi nghiên cứu và phương pháp nghiên cứu thực hiện đề tài.

1.1. Cơ sở khoa học và thực tiễn của đề tài

Vùng đất tây Nam Bộ là nơi mang một sắc thái cổ điển, ở đó chứa đựng tinh hoa về ẩm thực và văn hóa của dân tộc Việt Nam, nó cũng là vùng đất của những lễ hội truyền thống dân tộc. Trong những lần đi du lịch hoặc lễ hội,... cụ thể là ở miền Tây Nam Bộ. Chúng tôi nhận thấy còn rất nhiều người họ không biết về các món đặc sản ở vùng miền này, họ vào một quán ăn nào đó mà muốn gọi món đó mà không biết nó tên gì gây ra trở ngại. Chính vì thế, chúng tôi xây dựng một ứng dụng di động hỗ trợ quảng bá món ăn, đồng thời cũng hỗ trợ giúp nhận dạng món ăn đặc sản của từng tỉnh thành miền Tây nam Bộ, từ đó sẽ giúp cho những người khách du lịch, lễ hội,... có trải nghiệm tốt hơn về chuyến đi. Hiện tại, chúng tôi chưa tìm được các ứng dụng di động có tích hợp nhận diện món ăn Tây Nam Bộ nào. Do đó, việc xây dựng ứng dụng di động để giới thiệu các món ăn đặc trưng của Miền Tây Nam Bộ và cho phép người dùng nhận diện một số món ăn đặc trưng của Vùng là vô cùng cần thiết.

1.2. Những nghiên cứu liên quan

Hoashi và các cộng sự [1] đã sử dụng máy học để xây dựng website cho phép nhận diện 85 món Nhật bản với độ chính xác 62.5%. Kagaya và các cộng sự [2] sử dụng mô hình CNN để phát hiện và nhận diện món ăn. Yuzhen Lu [3] đề xuất sử dụng mạng Convolutional Neural Network (CNN) [4] với 3 lớp convolution pooling và 1 lớp fully connected để nhận dạng món ăn. Kết quả thực nghiệm trên tập dữ liệu gồm 5,822 hình ảnh thuộc 10 lớp món ăn cho thấy CNN đạt chính xác 90%, tốt hơn nhiều so với sử dụng SVM với độ chính xác 74%. Jeny và cộng sự [5] đề xuất mô hình FoNet dựa trên mạng deep residual neural network với 47 lớp CNN để nhận dạng các món ăn ở địa phương ở Bangladesh. Dựa trên tập dữ liệu các món ăn ở Bangladesh do tác giả xây dựng, FoNet vượt trội hơn Inception V3 [6] và MobileNet [7] với tỷ lệ chính xác lần lượt là 98,16%,

95,8% và 94,5%. Akti và cộng sự [8] chỉnh sửa Mobilenet-v2 [9] để nhận dạng các món ăn mở Trung Đông ở thời gian thực. Các tác giả xây dựng tập dữ liệu hình ảnh món ăn gồm 5.723 hình ảnh thuộc 27 lớp bằng cách thu thập từ Google và Instagram, đồng thời sử dụng các kỹ thuật tăng cường dữ liệu trên các lớp có dưới 100 hình ảnh. Thực nghiệm cho thấy sử dụng Mobilenet-v2 để nhận dạng món ăn đạt độ chính xác top-5 là 99,5%. Shen và cộng sự [10] đề xuất một hệ thống nhận dạng món ăn hoạt động theo mô hình client-server và ước tính lượng dinh dưỡng bằng cách tinh chỉnh mô hình Inception-V3 và Inception-V4. Tác giả thực hiện các thực nghiệm trên tập dữ liệu Food-101 của mình và đạt độ chính xác 85%. Zahisham và cộng sự [11] tinh chỉnh mô hình ResNet-50 [12] dựa trên Deep CNN (DCNN) để nhận các món ăn. Kết quả thử nghiệm dựa trên một số tập dữ liệu chuẩn cho thấy DCNN-ResNet có độ chính xác cao nhất là 41,08% vượt qua các mô hình CNN khác. DCNN-ResNet giúp nhận dạng món ăn đạt kết quả hơn cả sự kết hợp giữa CNN và SVM. Min và cộng sự [13] giải quyết việc nhận diện các món ăn bằng cách sử dụng mạng global-local attention xếp chồng lên nhau. Các đặc trưng trích xuất được kết hợp lại và biểu diễn thành vector đầu vào cho mô hình nhận dạng món ăn. Độ chính xác của mô hình trên tập ISIA Food-500 là 89,12%. Ngoài ra, Min và cộng sự [14] giới thiệu mạng Progressive Region Enhancement Network (PRENet) để rút trích đặc trưng của hình ảnh và mối liên hệ giữa các đối tượng nhằm cải thiện độ chính xác của mô hình nhận dạng hình ảnh. PRENet sử dụng RestNet50 làm mạng backbone. PRENet hoạt động tốt hơn tất cả các mô hình hiện có với độ chính xác top-5 là 97,33% trên tập dữ liệu Food2k và 98,71% trên tập dữ liệu ETH Food-101 [15]. Hiện tại, với sự phát triển mạnh mẽ của các mô hình học sâu, các mô hình khác nhau đã được sử dụng để phát hiện và nhận diện đối tượng có thể kể đến như sử dụng mô hình Convolutional Neural Network – CNN [16], Spatial Pyramid Pooling Networks – SPPNet [17], Faster RCNN [18], và YOLO [19]. Mô hình YOLO đạt kết quả vượt trội so với các mô hình học sâu khác.

Theo kiến thức của chúng tôi, không có quá nhiều nghiên cứu liên quan đến bài toán phát hiện và nhận diện món ăn nói chung và món ăn Việt Nam nói riêng. Ung và cộng sự [20] đã thu thập 8.903 hình ảnh món ăn Việt Nam thuộc 13 lớp và huấn luyện mô hình dựa trên CNN như AlexNet, GoogleNet, ResNet và InceptionResNet-v2 để nhận

điện món ăn Việt Nam. Kết quả thực nghiệm cho thấy mô hình InceptionResNet-v2 hoạt động tốt hơn các mô hình dựa trên CNN khác. Nguyen và cộng sự [21] đã xây dựng tập dữ liệu VinaFood21, bao gồm 13.950 hình ảnh của 21 món ăn phổ biến, để đánh giá nhận dạng món ăn Việt Nam. Mô hình CNN Efficientnet-B0 [22] được tinh chỉnh để nhận dạng món ăn Việt Nam và độ chính xác trung bình cao nhất là 74,81%. Tương tự, Do và cộng sự [23] xây dựng tập dữ liệu 30CNFoods gồm 25.136 hình ảnh về 30 món ăn phổ biến của Việt Nam. Các tác giả thực nghiệm với một số phương pháp học sâu và đạt được độ chính xác top-5 cao nhất là 97,07%.

Các trang web giới thiệu ẩm thực trong nước như hoangviettravel.vn¹, vinperal.com², hay pasgo.vn³ được giới thiệu một cách chung chung, kết hợp cung cấp các thông tin khác. Chúng tôi chưa tìm thấy ứng dụng di động có tích hợp nhận dạng món ăn đặc trưng vùng Tây Nam Bộ. Bài toán nhận diện và phân loại món ăn nói riêng, nhận diện và phân loại đối tượng nói chung có thể được giải quyết bằng mô hình học sâu. Do đó, trong nghiên cứu này, chúng tôi sẽ nghiên cứu tinh chỉnh các mô hình học sâu như YOLO, vision Transformer, và MobileNet để phát hiện và nhận diện một số món ăn đặc trưng của miền Tây Nam Bộ

1.3. Mục tiêu

Mục tiêu đề tài là xây dựng tập dữ liệu các món ăn đặc trưng của miền Tây Nam Bộ để huấn luyện mô hình. Sau đó, tích hợp vào ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ, và cho phép nhận diện một số món ăn đặc trưng của miền Tây Nam Bộ.

1.4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của đề tài là ứng dụng di động và các mô hình học sâu. Phạm vi nghiên cứu của đề tài là một số món ăn đặc trưng ở khu vực miền Tây Nam Bộ.

¹ https://hoangviettravel.vn/dac-san-mientay/#2_Mon_an_dac_san_mien_Tay_Nam_Bo

² <https://vinpearl.com/vi/ngat-ngay-voi-top-28-dac-san-mien-tay-ngon-kho-cuong>

³ <https://pasgo.vn/blog/top-15-dac-san-mien-tay-nam-bo-lamqua-duoc-chon-mua-nhieu-nhat-4000>

1.5. Phương pháp nghiên cứu

Để thực hiện nghiên cứu, chúng tôi sẽ tìm kiếm các công trình nghiên cứu đã được công bố, phân tích ưu nhược điểm của phương pháp, và đề xuất phương pháp phù hợp để giải quyết bài toán. Kế tiếp sẽ triển khai thực hiện huấn luyện và đánh giá mô hình. Cuối cùng, chúng tôi triển khai xây dựng mô hình thành ứng dụng di động.

1.6. Bộ cục

Bộ cục quyển báo cáo gồm 6 chương.

- Chương 1: Giới thiệu đề tài, trình bày nội dung mục tiêu đề tài, phạm vi và phương pháp nghiên cứu đề tài.
- Chương 2: Trình bày và phân tích mô hình học sâu như YOLO, Vision Transformer, MobileNet-v3.
- Chương 3: Trình bày phương pháp thực hiện tinh chỉnh các mô hình nhận dạng.
- Chương 4: Trình bày phương pháp xây dựng tập dữ liệu
- Chương 5: Thảo luận thực nghiệm và phương pháp xây dựng ứng dụng di động.
- Chương 6: Tổng kết kết quả đạt được của đề tài, thảo luận hạn chế và đề xuất hướng khắc phục.

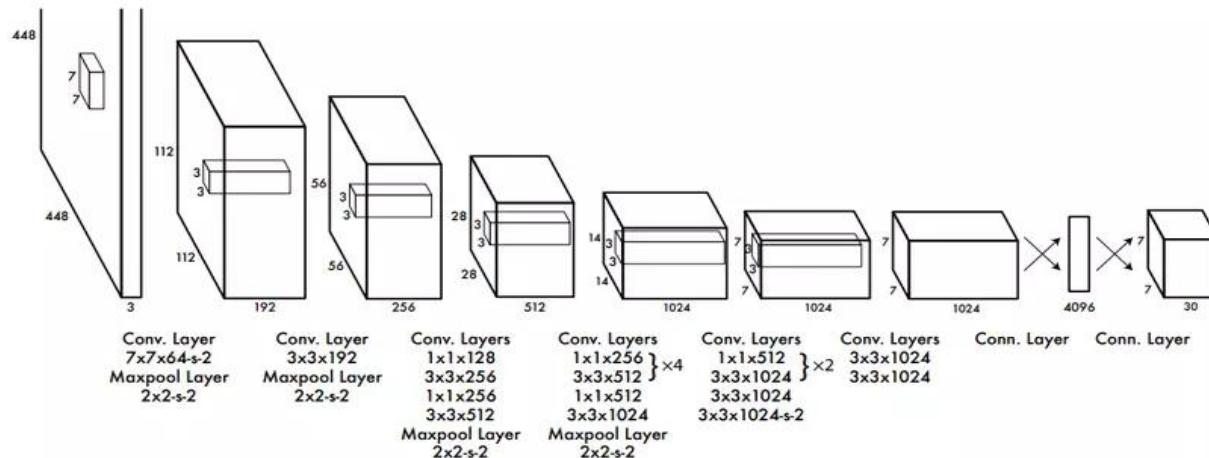
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

Chương 2 sẽ trình bày cơ sở lý thuyết về các mô hình sẽ được sử dụng để xây dựng hệ thống nhận dạng món ăn Tây Nam Bộ

2.1. Mô hình YOLO

2.1.1. Tổng quan YOLO

YOLO [19] là một thuật toán của mô hình mạng nơ ron tích chập CNN được phát triển cho việc phát hiện nhận dạng, phân loại đối tượng. YOLO được tạo ra từ việc kết hợp giữa các convolutional layers và connected layers. Trong đó các convolutional layers sẽ trích xuất ra các feature của ảnh, còn full-connected layers sẽ dự đoán ra xác suất đó và tọa độ của đối tượng. Về độ chính xác thì YOLO có thể không phải là thuật toán tốt nhất nhưng nó là thuật toán nhanh nhất (trong thời gian thực) so với các mô hình nhận dạng đối tượng khác. Kiến trúc mô hình YOLO được trình bày ở Hình 1.



Hình 1. Kiến trúc mô hình YOLO [19]

Tính đến thời điểm hiện tại YOLO đã có 7 phiên bản gồm v1,v2,v3,v4,v5,v6,v7 (Hình 2). Trong đó v5 là phiên bản theo nhóm nghiên cứu đánh giá ổn định, được cải thiện về tốc độ, độ chính xác và khắc phục được các nhược điểm của những phiên bản trước như: lỗi xác định vị trí của vật, ràng buộc về không gian bounding box. Đôi với 2 phiên bản mới nhất là v6 và v7 vừa ra mắt trong năm 2022 thì vẫn còn những thiếu sót chưa được tối ưu và đang được trong quá trình cải tiến.



Hình 2. YOLO timeline

2.1.2. Mô hình YOLOv5

YOLOv5 là phiên bản của dòng mô hình YOLO, nó phát hiện đối tượng bằng cách chia hình ảnh thành một hệ thống lưới. Mỗi ô trong lưới có nhiệm vụ phát hiện các đối tượng trong chính nó. YOLOv5 là mẫu YOLO đầu tiên được viết trong khuôn khổ PyTorch, nó nhẹ hơn và dễ sử dụng hơn. So với các khung phát hiện đối tượng khác, YOLOv5 cực kỳ dễ sử dụng cho một nhà phát triển triết khai các công nghệ thị giác máy tính vào một ứng dụng. YOLOv5 đề xuất người dùng 5 phiên bản chủ yếu, được trình bày ở Hình 3, là YOLOv5-n phiên bản nano, YOLOv5-s phiên bản nhỏ, YOLOv5-m phiên bản trung bình, YOLOv5-l phiên bản lớn, và YOLOv5-x phiên bản cực đại.



Hình 3. Các phiên bản của YOLOv5

Nguồn: <https://pythondig.com/repo/helmet-detection-using-YOLO-algorithm--pytorch>

Mô hình nhận dạng đối tượng được thiết kế để tạo ra đặc điểm từ hình ảnh đầu vào và sau đó cung cấp tính năng thông qua hệ thống dự đoán để vẽ các hộp xung quanh các

đối tượng và dự đoán các lớp của chúng. Mô hình YOLO là bộ nhận dạng đối tượng đầu tiên kết nối quy trình dự đoán các hộp giới hạn với các nhãn. Kiến trúc kết cấu của YOLOv5 phần lớn vẫn giữ nguyên so với bản V4, bao gồm như sau:

- Backbone: được sử dụng chủ yếu để trích xuất các phần quan trọng từ hình ảnh đầu vào đã cho. Trong YOLOv5, CSP (Cross Stage Partial Networks) được sử dụng làm xương sống để trích xuất tính năng thông tin từ hình ảnh đầu vào.
- Neck: được sử dụng tạo các kim tự tháp đặc trưng. Các kim tự tháp giúp các mô hình khái quát về tỷ lệ đối tượng, giúp xác định cùng một đối tượng với các kích thước và tỷ lệ khác nhau. Trong YOLOv5 sử dụng PANet để có được các kim tự tháp đặc trưng. Và chuyển nó cho Head để dự đoán.
- Head: chủ yếu được sử dụng để thực hiện phần phát hiện cuối cùng. Nó áp dụng các anchor boxes trên các tính năng và tạo ra các vector đầu ra cuối cùng với xác suất lớp, điểm đối tượng và các hộp giới hạn vật thể. Phần Head này YOLOv5 tương tự với các bản YOLOv3 và YOLOv4.

2.1.3. Mô hình YOLOv7

YOLOV7 [24] là mô hình phát hiện đối tượng thời gian thực nhanh nhất và chính xác nhất cho các tác vụ thị giác máy tính. YOLOV7 vượt qua mọi mô hình nhận dạng đối tượng về cả tốc độ và độ chính xác. YOLOv7 được huấn luyện trên tập dữ liệu COCO từ đầu mà không sử dụng bất kì pretrained nào. Kiến trúc YOLOV7 như sau:

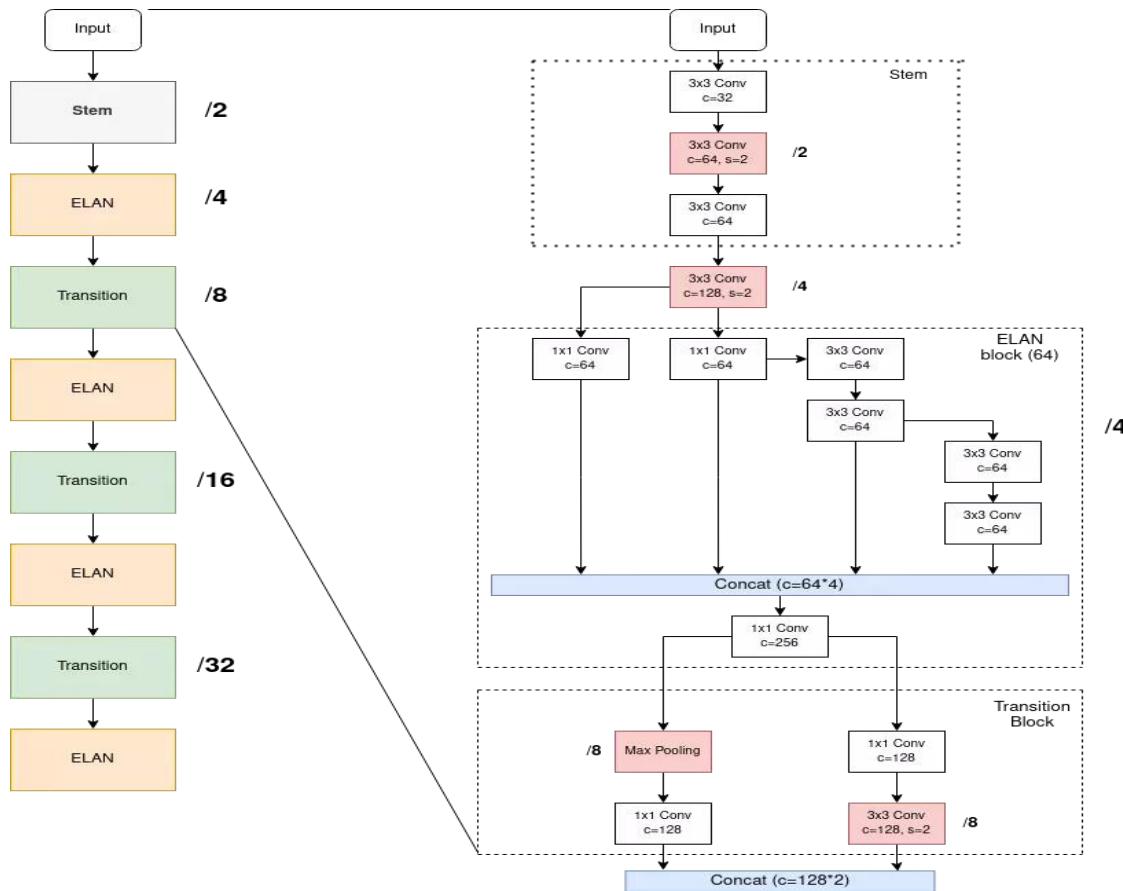
Backbone: được tạo từ các ELAN (Efficient Layer Aggregation Network) block. Một ELAN block gồm 3 phần: Cross Stage Partial (CSP), Computation Block và PointWiseConv. CSP hóa một block là việc tạo thêm một nhánh “cross stage partial”. Computation block là block chứa các lớp Conv được tính toán để sinh ra các feature mới thông qua các 3x3 Conv. Cuối cùng, các feature map được tổng hợp lại ở cuối sử dụng toán tử concatenate trên chiều channel như VoVNet, và đưa qua PointWiseConv (1x1 Conv).

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ

Transition Block: các ELAN Block được kết nối với nhau thông qua các Transition Block. Mỗi Transition Block là một lần giảm kích cỡ của mỗi feature map đi 2 lần.

Stem: trước khi tiến vào ELAN Block đầu tiên trong backbone, ảnh đầu vào sẽ đi qua Stem Block.

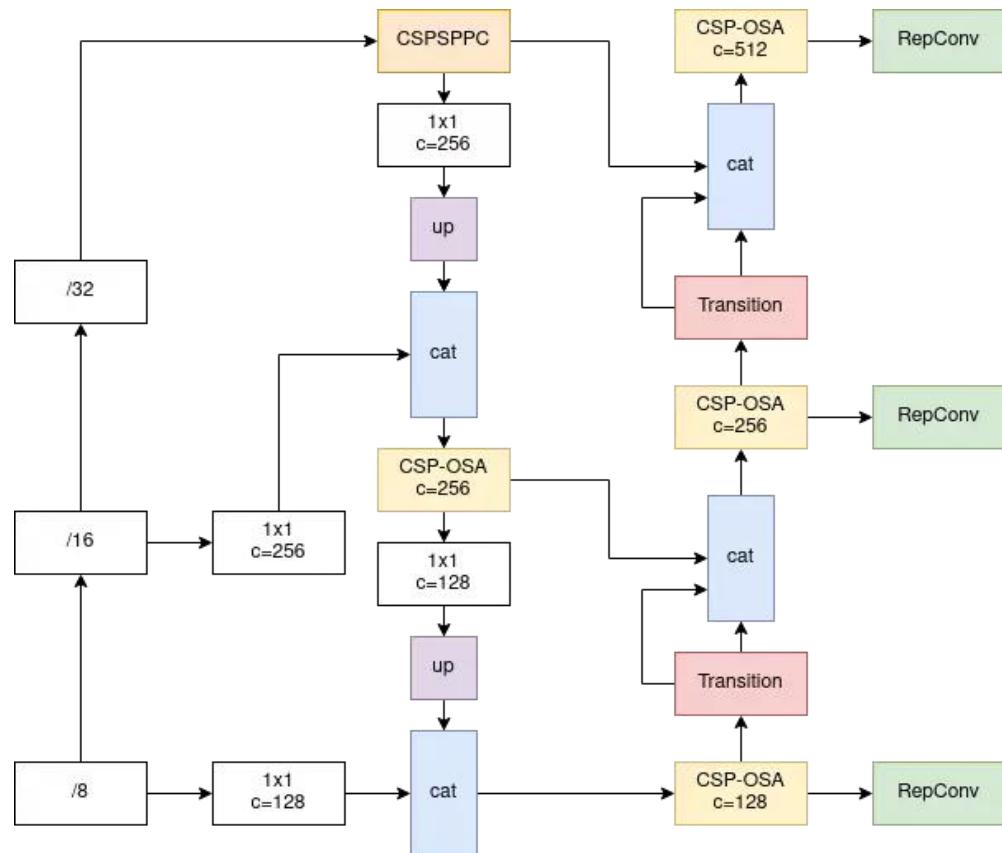
Backbone hoàn chỉnh của YOLOv7 là tập hợp các ELAN Block và các Transition block, như Hình 4.



Hình 4. Backbone của YOLOv7 [24]

Neck: YOLOv7 sử dụng CSPSPPC. Kiến trúc CSPSPPC được trình bày ở Hình 5. SPP lần đầu được áp dụng và YOLOv4, và được cải tiến thành SPPF trong YOLOv5. Còn trong YOLOv7, SPP tiếp tục được CSP hóa.

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ



Hình 5. Neck của YOLOv7[24]

RepConv: các feature maps từ các scale khác nhau sau đi đưa qua FPN hay PANet sẽ được xử lý thêm với 3×3 Conv ứng với từng scale. Còn trong Neck của YOLOv7, việc xử lý này sẽ được thực hiện bởi RepConv, một module với tốc độ của 3×3 Conv nhưng độ chính xác cao hơn rất nhiều. RepConv sử dụng kĩ thuật Re-param để có một module tốc độ cao và độ chính xác cũng cao.

Head: YOLOv7 chỉ đơn giản là YOLOR, sử dụng implicit knowledge

Cũng giống như các YOLO khác, v7 cũng có những phiên bản khác nhau của nó như: YOLOv7; YOLOv7x; YOLOv7d6; YOLOv7-p5; YOLOv7-p5 là nhóm model được train với size ảnh 640x640 và gồm 3 model là YOLOv7-tiny, YOLOv7 và YOLOv7x; YOLOv7-p6 là nhóm model được train với size ảnh 1280x1280 gồm 3 model là YOLOv7-w6, YOLOv7-e6, YOLOv7-d6; và YOLOv7-E6E.

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ

2.1.4. Output của YOLO

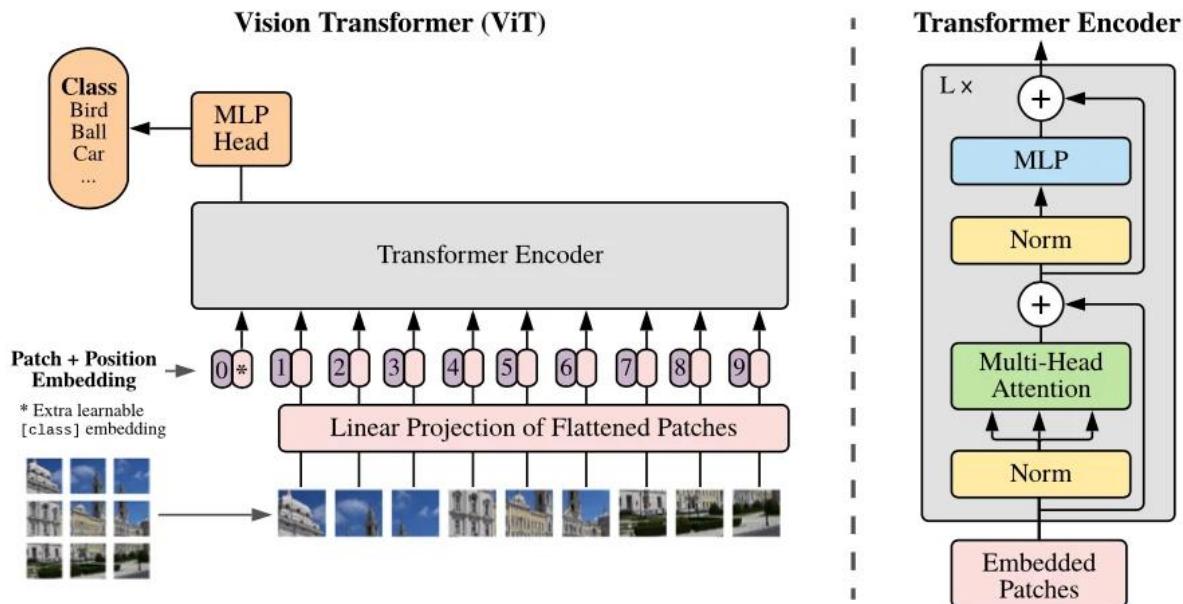
Output của mô hình YOLO là một véc tơ sê bao gồm các thành phần:

$$y^t = [p_0, < t_x, t_y, t_w, t_h, > < p_1, p_2, \dots, p_c, >] \quad (1)$$

Trong đó, p_0 là xác suất dự báo vật thể xuất hiện trong bounding box, $< t_x, t_y, t_w, t_h, >$ giúp xác định bounding box với t_x, t_y là tọa độ tâm và t_w, t_h là kích thước rộng dài của bounding box. $< p_1, p_2, \dots, p_c, >$ là vectơ phân phối xác suất dự báo của các lớp.

2.2. Mô hình Vision Transformer

Mô hình Vision Transformer (ViT) [25] dựa trên cơ chế attention và được áp dụng cho bài toán phân loại ảnh. Mô hình ViT dựa trên mô hình Transformer với kiến trúc Encoder-Decoder. Tuy nhiên, ViT chỉ sử dụng khối Encoder của mô hình Transformer. Kiến trúc của mô hình như Hình 6.



Hình 6. Kiến trúc mô hình Vision Transformer [25]

Patch Embedding: ViT sử dụng cơ chế Attention và giữ nguyên các khối Encoder của mô hình Transformer. Vấn đề ở đây là làm thế nào để đưa một hình ảnh vào khối Encoder trong khi Transformer sử dụng dữ liệu đầu vào là một câu. Dosovitskiy và các

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

công sự [25] xem hình ảnh là tập hợp những từ ngữ có kích thước 16×16 . Cụ thể, ViT sẽ tìm cách biến đổi hình ảnh về hình dạng của một câu để đưa vào khối Encoder. Với một hình ảnh đầu vào 3D Image: $(x) \in R^{H \times W \times C}$, mô hình sẽ tạo ra N flattened 2D patch

$$\text{Patch Image: } (x)_p \in R^{N \times (P^2 \times C)} \quad (2)$$

Với $N = (H \times W)/P^2$ và P là patch size. Khi đó, mỗi patch có kích thước $P^2 \times C$ tương ứng với một từ và N được xem như độ dài của câu trong mô hình Transformer.

Class token: sau khi thu được chuỗi Patch Embedding, mô hình sẽ gắn thêm class token vào đầu của nó để đánh dấu hình ảnh thuộc phân lớp nào. Class token là ma trận có kích thước $1 \times (P^2 \times C)$ bằng với kích thước của 2D patch. Giờ đây độ dài của chuỗi Patch Embedding là $N + 1$.

Positional Embedding: trong bài toán NLP, dữ liệu đầu vào là một câu. Nếu ở RNN, sau khi tách câu thành từng từ và các từ này được đưa vào khối Encoder một cách tuần tự khi ở Transformer, các từ này được đưa vào đồng loạt. Vì thế, việc đánh dấu vị trí của từng từ trong câu là rất quan trọng, vì chỉ cần hoán đổi vị trí một cặp hoặc nhiều từ thì ý nghĩa của câu sẽ thay đổi. Và việc đó cũng không ngoại lệ đối với ViT. Thủ xét ví dụ ở Hình 7:



Hình 7. Cùng một ảnh đầu vào và cách chia patch nhưng thứ tự các patch khác nhau

Nếu không đánh dấu vị trí của từng patch thì 2 ảnh ở trên hoàn toàn không có sự khác biệt.

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ

Trong mô hình ViT, Positional Embedding có nhiệm vụ chứa thông tin về vị trí của các patch trong ảnh (spatial information). Ma trận Positional Embedding có kích thước model được xác định theo công thức:

$$PE_{(pos,2i)} = \sin\left(\frac{POS}{1000^{\frac{2i}{d_{model}}}}\right) \quad (3)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{POS}{1000^{\frac{2i+1}{d_{model}}}}\right) \quad (4)$$

Trong đó, pos là vị trí của patch và i là vị trí phần tử trong Positional Embedding.

Sau đó, ViT sẽ cộng Patch Embedding và Positional Embedding để thu được một ma trận tạm gọi là Combined Embedding.

❖ Khối Transformer Encoder và Multi-Layer Perceptron Head

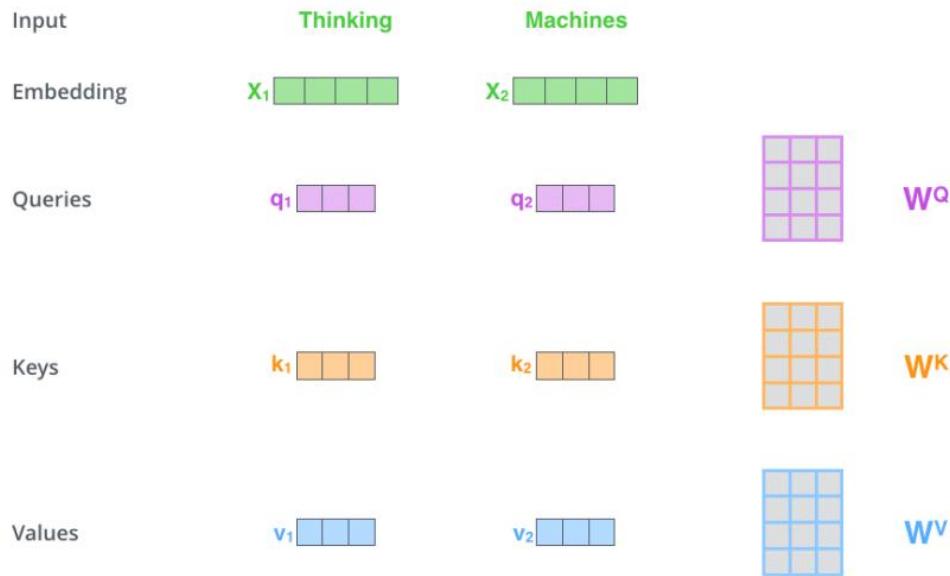
Thành phần chính của Transformer Encoder là Multi-Head Self-Attention (MSA) và Multi-Layer Perceptron(MLP). MSA chia input thành nhiều đầu để mỗi đầu có thể học cách tự chú ý ở mức độ khác nhau. MSA gồm 12 đầu self-attention. Output của các đầu sau đó được nối lại và đưa qua MLP. Ngoài ra các lớp Layernorm (LN) đứng trước và kết nối dư(residual connection) đứng sau mỗi khối MSA và khối MLP. Tác giả cho rằng làm vậy sẽ giúp mô hình huấn luyện dễ dàng và hội tụ nhanh hơn. Cuối cùng là MLP có chứa 2 lớp với GELU phi tuyến.

Phép toán tự chú ý với mục tiêu là thể hiện mức độ liên quan của một patch với tất cả các patch trong ảnh (kể cả với chính patch đó). Để đơn giản, ta xét phép tính tự chú ý trên vector. Việc đầu tiên là tạo ra 3 vector query (q), key (k) và value (v) từ vector đầu vào. Bộ ba Q, K, V được tính theo công thức:

$$q = x * WQ \quad k = x * WK \quad v = x * WV \quad (5)$$

Trong đó x là input vector, W^Q, W^K, W^V là những ma trận được khởi tạo ngẫu nhiên ban đầu và có thể học được.

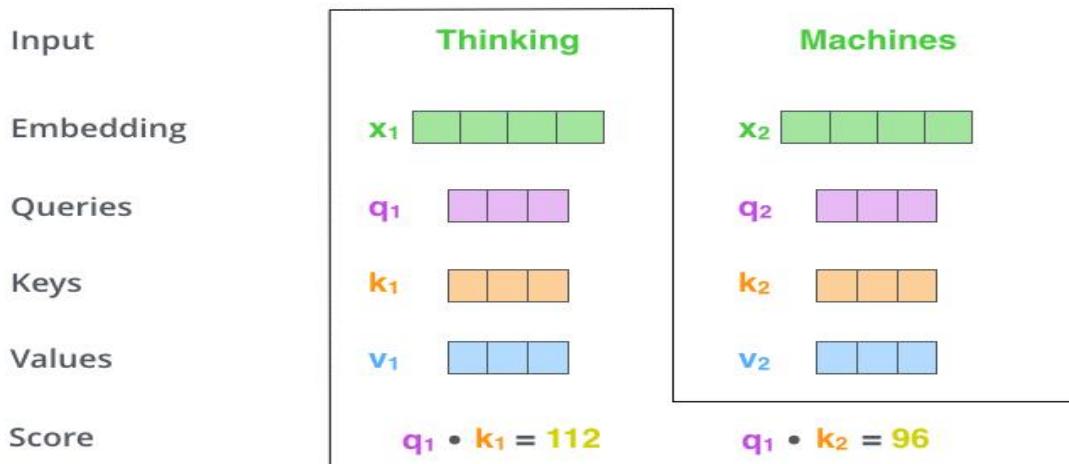
Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ



Hình 8. Khởi tạo các vector q, k, v

Nguồn: <https://jalammar.github.io/illustrated-transformer/>

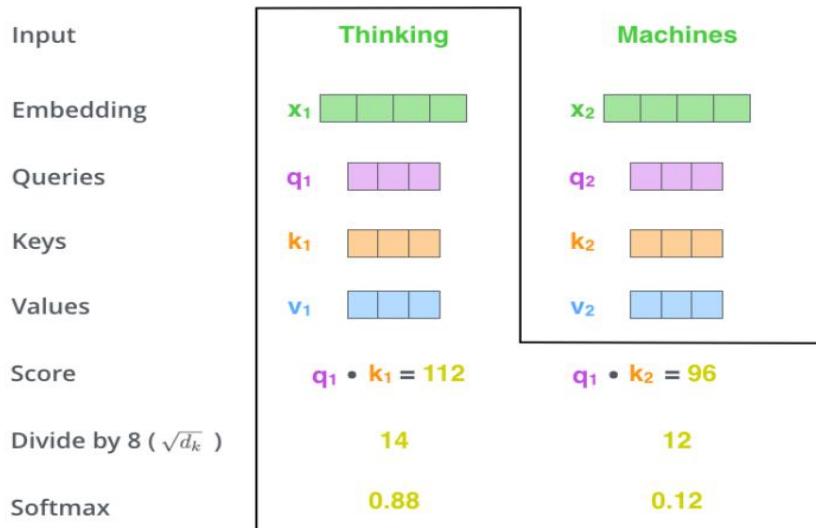
Giả sử ta có 2 patch tương ứng với 2 vector embedding là x_1 và x_2 . Để tính điểm chú ý cho x_1 ta thực hiện như sau: Đầu tiên, nhân vô hướng vector q_1 lần lượt với tất cả các vector k đang có ta thu được các score như Hình 9.



Hình 9. Nhân lần lượt các vector q với vector k

Nguồn: <https://jalammar.github.io/illustrated-transformer/>

Tiếp theo, chia các score chia căn bậc 2 kích thước của vector k . Giả sử kích thước vector k là 64. Sau đó cho kết quả qua hàm softmax như Hình 10.



Hình 10. Chia score cho căn bậc 2 kích thước vector k rồi đưa qua hàm softmax

Nguồn: <https://jalammar.github.io/illustrated-transformer/>

Các giá trị softmax trên được tính như sau:

$$\text{sum} = e^{14} + e^{12} \quad (6)$$

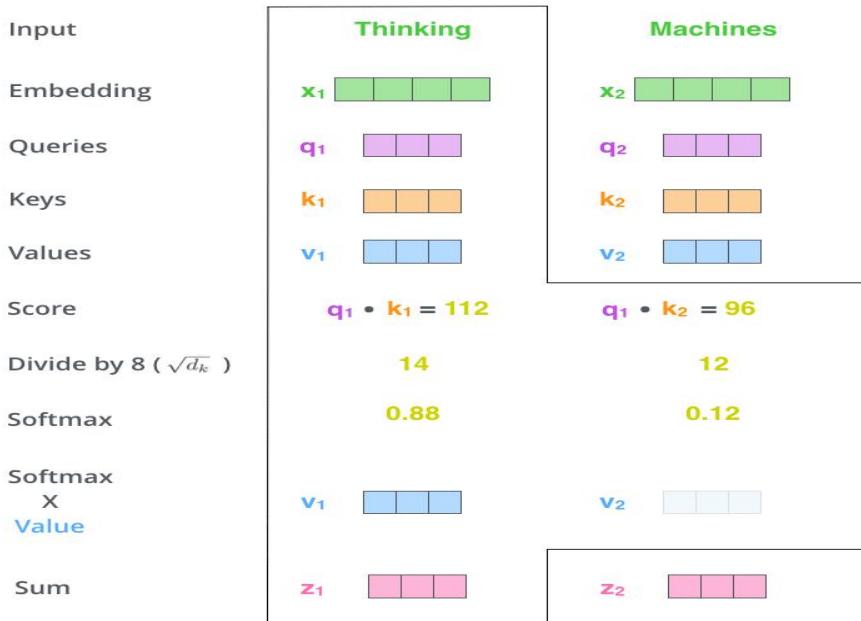
$$\text{Softmax}(x_1, x_1) = (e^{14})/\text{sum} \quad (7)$$

$$\text{Softmax}(x_1, x_2) = (e^{12})/\text{sum} \quad (8)$$

Giá trị softmax của một patch tại một patch khác thể hiện mức độ liên quan của 2 patch đó. Đương nhiên một patch sẽ liên quan đến chính nó nhiều nhất, nhưng chỉ số này hữu ích khi đánh giá độ liên quan với các patch còn lại.

Ké tiếp nhận vector v với giá trị softmax. Mục đích của việc này là giữ lại các patch liên quan và loại bỏ các patch không liên quan, vì tích của phép nhân tỉ lệ thuận với điểm softmax, mà điểm softmax lại tỉ lệ thuận với độ liên quan.

Cuối cùng tính tổng các tích vừa nhận được ở bước trên. Đây cũng là output của lớp Self-Attention.



Hình 11. Nhân giá trị softmax với vector v sau đó cộng các vector kết quả lại

Nguồn: <https://jalammar.github.io/illustrated-transformer/>

Sau khi có được output từ MSA, chúng được đưa vào các đầu residual connection để tạo ra output cuối cùng trước khi qua lớp LN và đi vào khối MLP. Output của MLP lại tiếp tục đưa vào các đầu residual connection để tạo ra output hoàn chỉnh của khối Transformer Encoder. Cuối cùng, output của khối Transformer Encoder được đưa vào Multi-Layer Perceptron Head để thực hiện dự đoán.

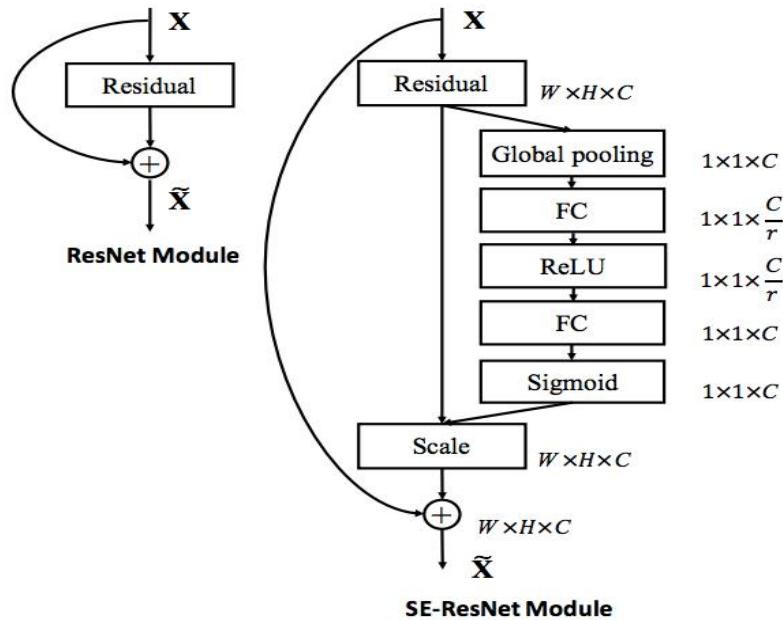
2.3. Mô hình MobileNet-v3

Mô hình MobileNet là mô hình sử dụng cách tích chập đặc biệt có tên là tích chập sâu phân tách (Depth Wise Separable Convolution) để giảm kích thước của mô hình và giảm độ phức tạp tính toán. Do đó, mô hình này rất nhẹ và hoạt động ổn định trên các ứng dụng di động và các thiết bị nhúng.

Kể từ khi ra đời, MobileNet-V3 [26] là một trong những kiến trúc được ưa chuộng nhất khi phát triển các ứng dụng AI trong thị giác máy tính. MobileNet-V3 có một số điểm cải tiến so với các phiên bản thấp hơn của nó giúp cho nó có độ chính xác cao hơn, số lượng tham số và số lượng các phép tính ít hơn. Mobilenet-V3 được điều chỉnh cho phù hợp với CPU của điện thoại. Hiện nay có hai mô hình mới của MobileNet cho ra độ

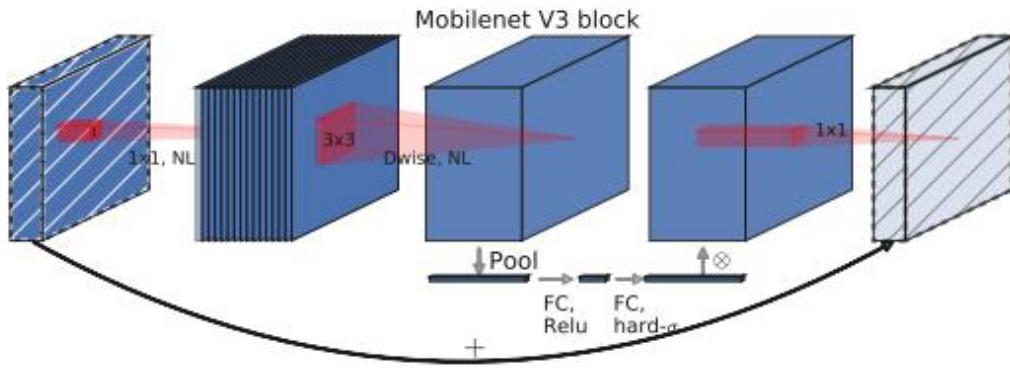
chính xác cao là MobileNet-V3-Large và MobileNet-V3-Small. Kết quả đạt được của mô hình MobileNet-V3-Large có chính xác hơn 3.2% về phân loại hình ảnh trong ImageNet trong khi độ trễ giảm 20% so với MobileNet-V2. MobileNet-V3-Small có độ chính xác hơn 6.6% so với MobileNet-V2 và có độ trễ tương đương. MobileNet-V3-Large nhận dạng nhanh hơn 25% so với MobileNet-V2 khi có cùng độ chính xác trên tập nhận dạng COCO. MobileNet-V3-Large Lite Reduced Atrous Spatial Pyramid Pooling(LR-ASPP) nhanh hơn 34% so với MobileNet-V2 R-ASPP khi có cùng độ chính xác tương đối ở phân đoạn Cityscapes [26].

MobileNet-v3 thêm Squeeze and Excitation (SE) vào block Residual để tạo thành một kiến trúc có độ chính xác cao hơn, như minh họa ở Hình 12.



Hình 12. Mô hình SE-ResNet Module[28]

SE-ResNet áp dụng thêm một nhánh Global pooling có tác dụng ghi nhận bối cảnh của toàn bộ layer trước đó. Kết quả sau cùng ở nhánh này ta thu được một véc tơ global context được dùng để scale đầu vào X . Tương tự như vậy SE được tích hợp vào kiến trúc của một residual block trong MobileNet-v3 như Hình 13 sau:



Hình 13. MobileNet-V3 block [26]

Tại layer thứ 3 có một nhánh Squeeze and Excitation có kích thước (*width* x *height*) bằng 1 x 1 có tác dụng tổng hợp global context. Nhánh này lần lượt đi qua các biến đổi FC → Relu → FC → hard sigmoid (FC là fully connected layer). Cuối cùng được nhân trực tiếp vào nhánh input để scale input theo global context. Các kiến trúc còn lại hoàn toàn giữ nguyên như MobileNet-V2.

2.4. Phương pháp đánh giá precision, recall, AP và mAP

Khi xây dựng mô hình phân loại chúng ta sẽ muốn biết một cách khái quát tỷ lệ các trường hợp được dự báo đúng trên tổng số các trường hợp là bao nhiêu. Tỷ lệ đó được gọi là độ chính xác. Độ chính xác giúp ta đánh giá hiệu quả dự báo của mô hình trên một bộ dữ liệu. Độ chính xác càng cao thì mô hình của chúng ta càng chuẩn xác.

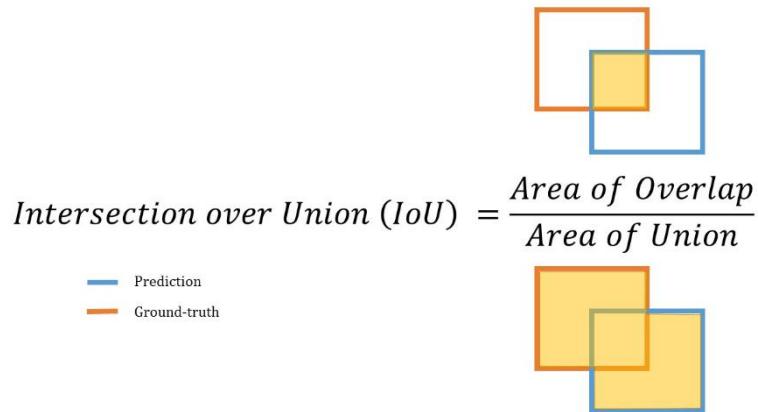
- TP (True Positive): Tổng số trường hợp dự báo khớp Positive.
- TN (True Negative): Tổng số trường hợp dự báo khớp Negative.
- FP (False Positive): Tổng số trường hợp dự báo các quan sát thuộc nhãn Negative thành Positive.
- FN (False Negative): Tổng số trường hợp dự báo các quan sát thuộc nhãn Positive thành Negative.
- Precision(độ chính xác) cho chúng ta biết rằng trong số các kết quả được phân loại tích cực trong mô hình, có bao nhiêu kết quả tích cực thực sự.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (9)$$

- Recall(thu hồi) chúng tôi biết có bao nhiêu điểm tích cực thực sự đã được mô hình của chúng ta thu hồi hoặc tìm thấy.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (10)$$

- Intersection over Union (IoU), hay còn gọi là Jaccard Index, là phương pháp được sử dụng để đo độ chính xác của các mô hình nhận diện đối tượng trên một tập dữ liệu cụ thể.



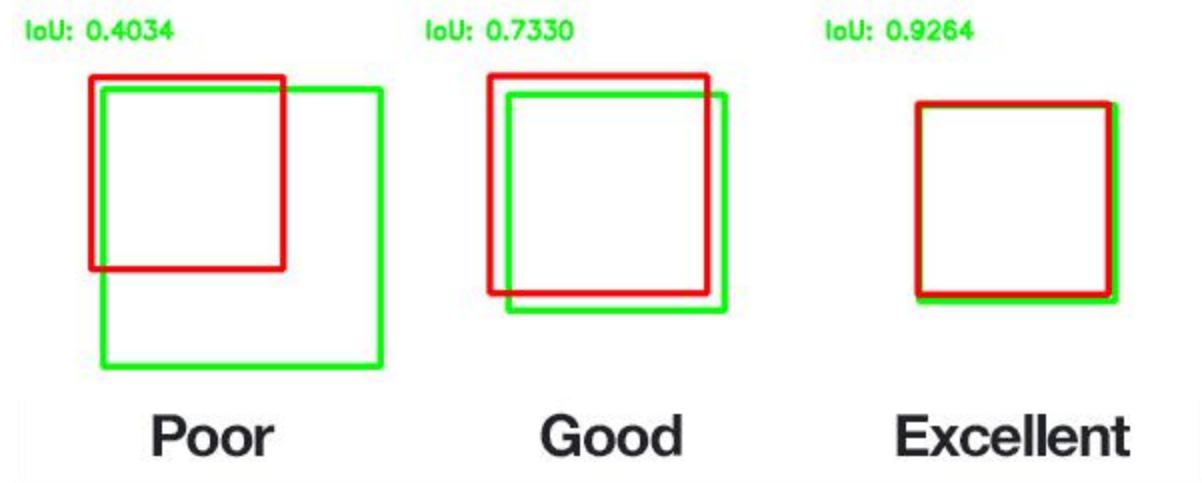
Hình 14. Công thức tính IoU

Nguồn: <https://www.kaggle.com/competitions/global-wheat-detection/overview/evaluation>

Trong đó:

- The area of overlap: diện tích phần giao nhau giữa đường bao thực và đường bao dự đoán.
- The area of union: diện tích phần hợp giữa đường bao dự đoán và đường bao thực tế.
- IoU được tính bằng cách chia area of overlap cho area of union, hay nói cách khác, IoU là thương của area of overlap chia cho area of union.

Tỷ lệ chính xác được xếp ở mức khá tốt khi IoU cho ra kết quả lớn hơn 0.5. Đồng nghĩa với việc đối tượng được xem là nhận dạng đúng.



Hình 15. Giá trị IoU cho bounding box

Nguồn: <https://intrepidgeeks.com/tutorial/yolo-object-detection>

- AP(Average Precision) là độ chính xác trung bình của từng lớp

$$AP = \int_{r=0}^1 p(r) dr \quad (11)$$

- mAP(Mean Average Precision) là độ chính xác trung bình của mô hình được tính theo công thức sau:

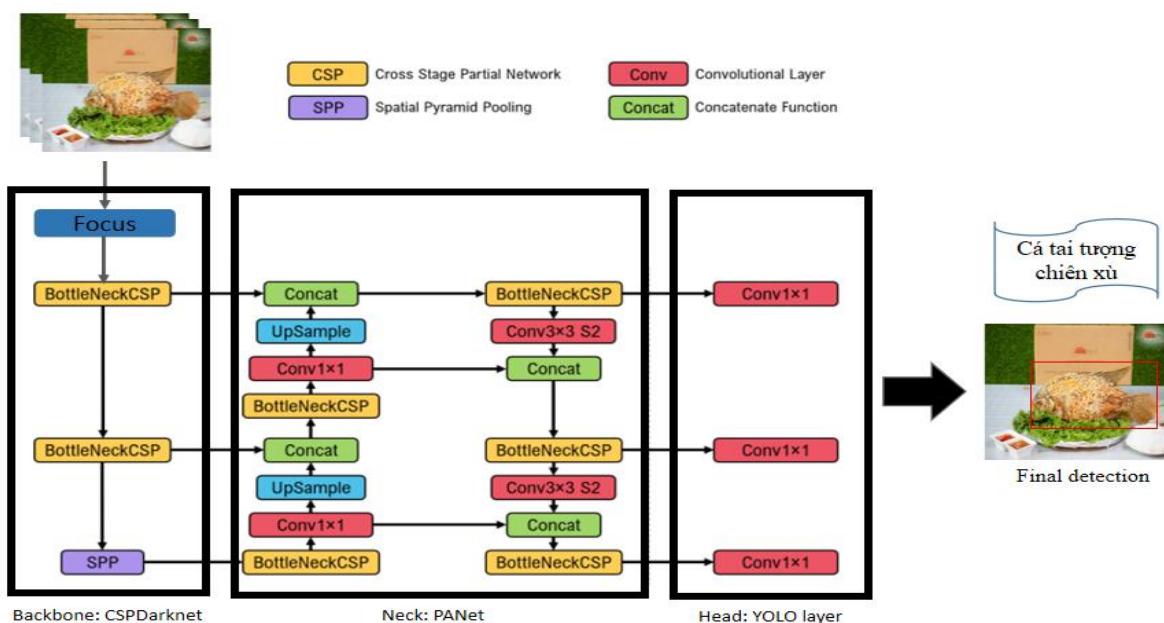
$$MAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (12)$$

CHƯƠNG 3. PHƯƠNG PHÁP THỰC HIỆN

Trong chương này, chúng tôi trình bày phương pháp thực hiện tinh chỉnh các mô hình YOLO, Vision Transformer và MobileNet để nhận diện món ăn Tây Nam Bộ.

3.1. Mô hình YOLO

Hình 16 trình bày mô hình YOLOv5 được tinh chỉnh cho bài toán nhận diện món ăn, dựa trên kiến trúc mô hình YOLOv5 của Renjie và các cộng sự [29]. Chúng tôi thực hiện thử nghiệm với YOLOv5 được cung cấp bởi Ultralytics⁴ và YOLOv7 được hỗ trợ bởi WongKinYiu⁵. Các mô hình YOLO được huấn luyện với batch size là 16, và 150 epoch. Các thông số khác, chúng tôi sử dụng các giá trị tương tự như đề xuất của Glenn Jocher⁶: initial learning rate là 0.001, final learning rate là 0,1, optimizer weight decay là 0,0005, và Adam optimizer.



Hình 16. Mô hình YOLO5 cho bài toán nhận diện món ăn Tây Nam Bộ.

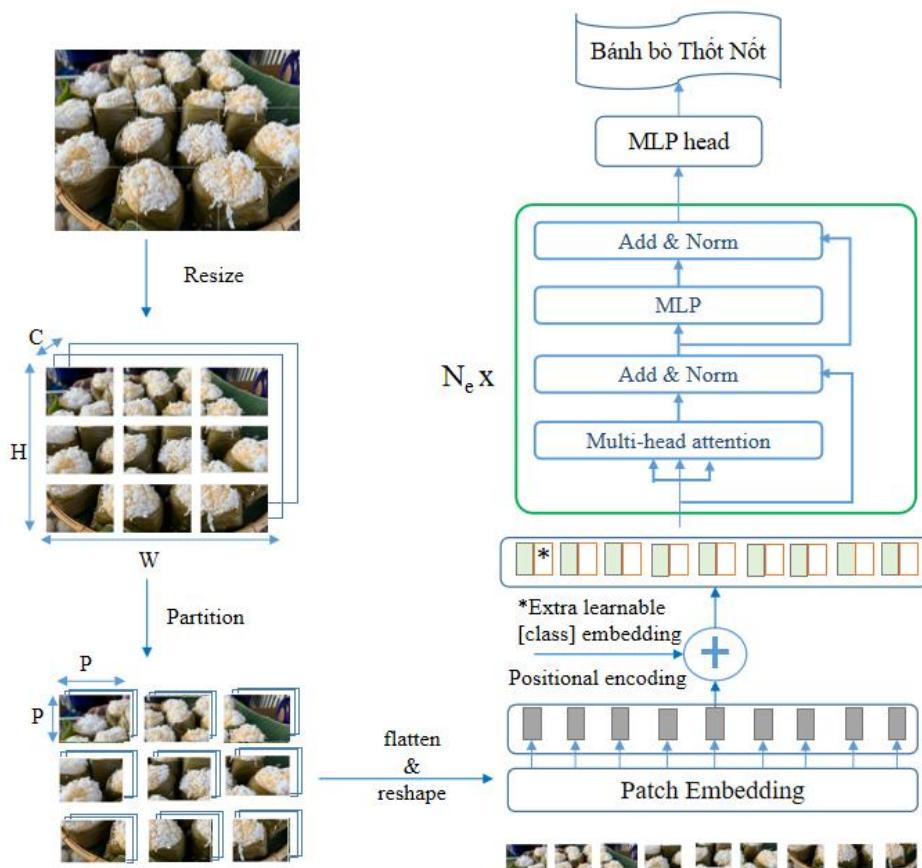
⁴ <https://ultralytics.com/yolov5>

⁵ <https://github.com/WongKinYiu/yolov7>

⁶ <https://github.com/ultralytics/yolov5>

3.2. Mô hình Vision Transformer

Trong nghiên cứu này, chúng tôi sử dụng mô hình ViT do Kaggle⁷ cung cấp, hình ảnh có kích thước cố định là 16x16. Mô hình được huấn luyện với patch size là 16, số epoch là 100 và Adam optimizer. Chúng tôi sử dụng mô hình ViT-base với số layers, hidden size, MLP size, and multi-head attention tương ứng là 8, 512, 2048, và 8. Hình 17 minh họa mô hình Vision Transformer được chỉnh sửa để nhận dạng các món ăn.

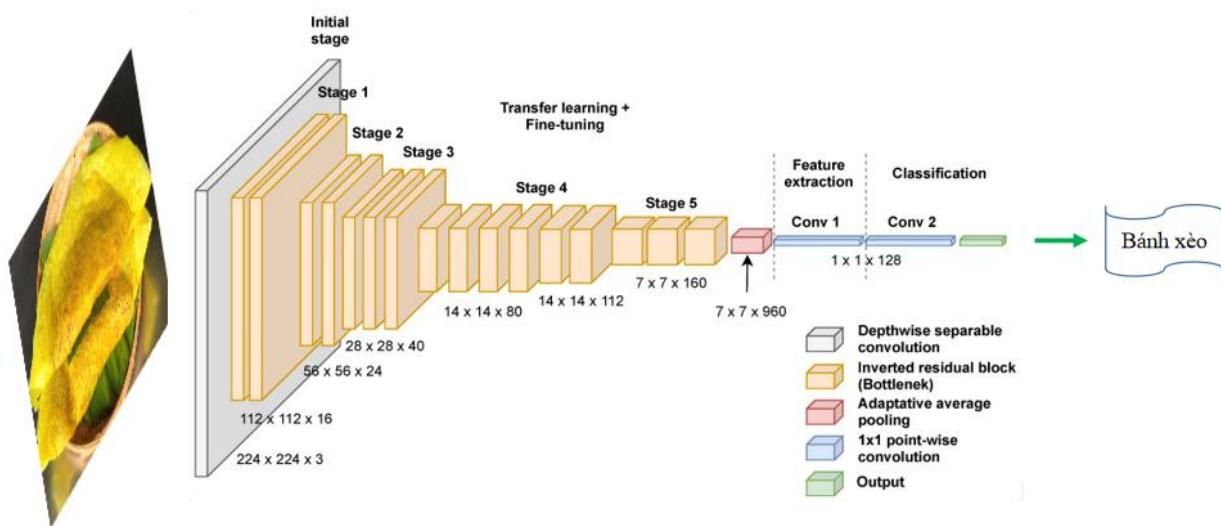


Hình 17. Mô hình Vision Transformer cho bài toán nhận diện món ăn Tây Nam Bộ.

⁷ <https://www.kaggle.com/>

3.3. Mô hình MobileNet-v3

Hình 18 minh họa mô hình MobileNet-v3 đã qua chỉnh sửa để nhận dạng các món ăn. Chúng tôi thực hiện huấn luyện mô hình với mô hình MobileNetV3-large, cung cấp bởi Kaggle, với patch size là 32, số epoch là 100, và các thông số khác theo đề xuất của tác giả⁸ như learning rate là 0,16, và RMSProp optimizer là 0,9 momentum.



Hình 18. Mô hình MobileNet-V3 cho bài toán nhận diện món ăn Tây Nam Bộ.

⁸ <https://github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet>

CHƯƠNG 4. XÂY DỰNG TẬP DỮ LIỆU

4.1. Thu thập dữ liệu

Chúng tôi xây dựng bộ dữ liệu 7.201 hình ảnh (với 4.179 ảnh được tạo ra sau khi dùng các phương pháp augmentation và 3.022 ảnh được thu thập) bao gồm 50 món ăn: Bánh Bò Chăm, Bánh Bò Thốt Nốt, Bánh Cam, Bánh Canh Ghẹ, Bánh Còng, Bánh Cống, Bánh Chuối, Bánh Đúc Lá Dừa, Bánh Đúc Mặn, Bánh Khot, Bánh Lá Dừa, Bánh Ông Lá Dừa, Bánh Pía, Bánh Tai Yến, Bánh Tầm Bì, Bánh Tầm Cay, Bánh Tầm Khoai Mì, Bánh Tét Lá Cảm, Bánh Xèo, Bún Bò Cay, Bún Cá, Bún Kèn Hà Tiên, Bún Quậy, Cá Kèo Nướng Muối Ót, Cá Kho Tộ, Cá Lóc Nướng, Cá Tai Tượng Chiên Xù, Canh Chua, Cơm Cháy Kho Quẹt, Cơm Gói Lá Sen, Cơm Tấm Long Xuyên, Chè Bưởi, Chuối Nép Nướng, Dừa Sáp, Duông Dừa, Gà Đốt Ô Thum, Gỏi Ba Khía, Gỏi Cá Trích, Gỏi Cuốn, Gỏi Sầu Đâu, Hủ Tiếu Mỹ Tho, Kẹo Dừa, Lẩu Gà Chanh Ót, Lẩu Mắm, Nem Lai Vung, Óc Len Xào Dừa, Óc Nướng Tiêu Xanh, Thốt Nốt, Vịt Nấu Chao, Xôi Phòng. Dữ liệu hình ảnh thu thập được đa dạng, với các hình dạng khác nhau của các món ăn và các màu sắc. Nguồn dữ liệu chính trong nghiên cứu được thu thập từ Google và tự chụp. Một số hình ảnh món ăn được minh họa ở Hình 19.



Hình 19. Ví dụ một số hình ảnh từ tập dữ liệu

Sau đó, phân loại và gán nhãn thủ công với sự hỗ trợ của phần mềm Roboflow⁹. Chủ thích và hình ảnh ví dụ được hiển thị ở Hình 20.



Hình 20. Gán nhãn thủ công với Roboflow

4.2. Tiền xử lý dữ liệu

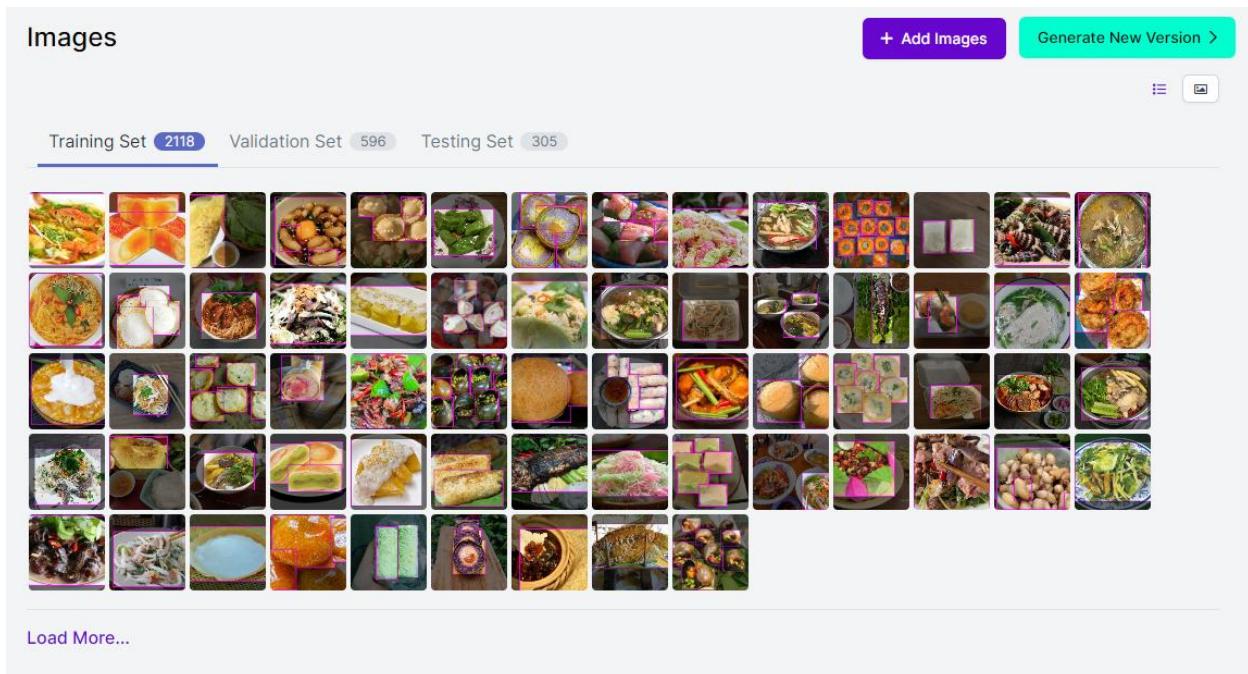
4.2.1. Tách tập dữ liệu

Tập dữ liệu tiền xử lý có 3.022 hình ảnh được chia thành 3 phần: 70% tổng số hình ảnh được sử dụng để huấn luyện mô hình, 20% tổng số hình ảnh để xác thực mô

⁹ <https://roboflow.com/>

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

hình train và 10% tổng số hình ảnh đánh giá mô hình. Hình 21 trình bày dữ liệu sau khi được phân chia.



Hình 21. Dữ liệu sau khi được phân chia

4.2.2. Tăng cường dữ liệu

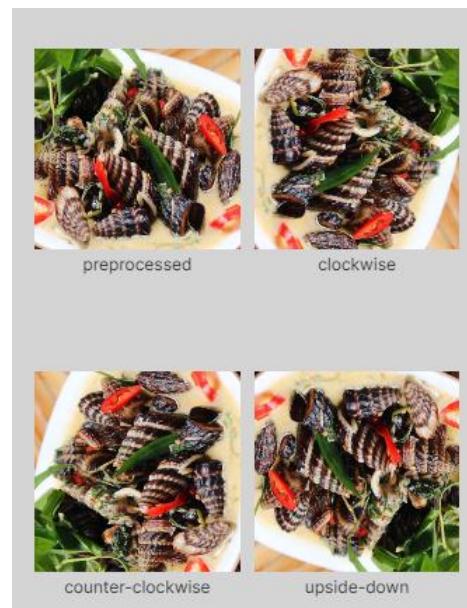
Do dữ liệu huấn luyện nhỏ, nên việc tăng cường dữ liệu là một kỹ thuật phổ biến để làm cho mô hình tổng quát hơn. Một số kỹ thuật tăng cường dữ liệu được sử dụng:

- Flip (hình 22): lật ảnh theo chiều ngang và chiều dọc



Hình 22. Minh họa kỹ thuật Flip

- 90° Rotate (Hình 23): Thêm góc xoay 90° để giúp mô hình không phụ thuộc với hướng máy ảnh: theo chiều kim đồng hồ, ngược chiều kim đồng hồ, lật ngược



Hình 23. Minh họa kỹ thuật 90° Rotate

- Blur (Hình 24): thêm hiệu ứng Blur để giúp mô hình linh hoạt hơn với tiêu điểm máy ảnh.



Hình 24. Minh họa kỹ thuật Blur

Bảng 1 trình bày số lượng ảnh trong dataset thu thập được và sau quá trình tăng cường dữ liệu.

STT	Tên đối tượng	Số lượng ảnh	Số lượng ảnh sau khi tăng cường dữ liệu
1	Bánh Bò Chăm	51	125
2	Bánh Bò Thốt Nốt	82	196
3	Bánh Cam	51	125
4	Bánh Canh Ghẹ	75	179
5	Bánh Còng	55	131
6	Bánh Cóng	50	120
7	Bánh Chuối	61	126
8	Bánh Đúc Lá Dứa	64	152
9	Bánh Đúc Mặn	51	121
10	Bánh Khọt	73	175
11	Bánh Lá Dừa	62	148
12	Bánh Ống Lá Dứa	50	120
13	Bánh Pía	61	145
14	Bánh Tai Yến	51	121
15	Bánh Tăm Bì	72	171
16	Bánh Tầm Cay	50	120

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

STT	Tên đồ vật	Số lượng ảnh	Số lượng ảnh sau khi tăng cường dữ liệu
17	Bánh Tăm Khoai Mì	60	144
18	Bánh Tét Lá Cầm	50	120
19	Bánh Xèo	64	152
20	Bún Bò Cay	66	158
21	Bún Cá	82	196
22	Bún Kèn Hà Tiên	55	131
23	Bún Quậy	66	155
24	Cá Kèo Nướng Muối Ót	55	129
25	Cá Kho Tộ	81	193
26	Cá Lóc Nướng	60	144
27	Cá Tai Tượng Chiên Xù	50	120
28	Canh Chua	88	210
29	Cơm Cháy Kho Quẹt	53	127
30	Cơm Gói Lá Sen	56	134
31	Cơm Tấm Long Xuyên	53	123
32	Chè Bưởi	60	144
33	Chuối Nép Nướng	60	144
34	Dừa Sáp	50	122
35	Đuông Dừa	50	122
36	Gà Đốt Ô Thum	75	179
37	Gỏi Ba Khía	54	128
38	Gỏi Cá Trích	65	155
39	Gỏi Cuốn	98	234
40	Gỏi Sầu Đâu	52	123
41	Hủ Tiếu Mỹ Tho	57	135
42	Kẹo Dừa	57	135
43	Lẩu Gà Chanh Ót	66	158
44	Lẩu Mắm	60	144
45	Nem Lai Vung	55	129
46	Óc Len Xào Dừa	51	122
47	Óc Nướng Tiêu Xanh	50	122
48	Thốt Nốt	51	122
49	Vịt Nấu Chao	63	151

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

STT	Tên đối tượng	Số lượng ảnh	Số lượng ảnh sau khi tăng cường dữ liệu
50	Xôi Phòng	50	121
	Tổng	3022	7201

Bảng 1. Số lượng hình ảnh trong dataset sau khi sử dụng các kỹ thuật tăng cường dữ liệu

4.2.3. Chuyển đổi dữ liệu sang định dạng YOLO

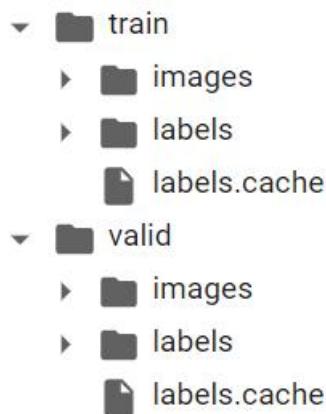
Để huấn luyện mô hình YOLO, dữ liệu cần được chuyển đổi sang định dạng YOLO chứa mỗi tệp hình ảnh và tệp văn bản cùng tên. Tệp văn bản chứa thông tin về các chủ thích và có cùng tên với các hình ảnh tương ứng của nó. Tệp chứa mỗi dòng cho mỗi đối tượng có trong hình ảnh. Ví dụ: nếu có 3 đối tượng tệp sẽ chứa 3 dòng. Mỗi dòng chứa thông tin chi tiết về các đối tượng riêng lẻ. Các tham số hoặc chi tiết các nhãn, vị trí x và y cũng như chiều cao và chiều rộng của đối tượng. Các đối tượng này được chuẩn hóa trong khoảng 0-1. Nhờ có Roboflow, việc chuyển đổi dữ liệu sang định dạng YOLO theo mong muốn rất nhanh chóng.

CHƯƠNG 5. THỰC NGHIỆM VÀ ỨNG DỤNG

Chương này sẽ trình bày cách cài đặt, thiết lập môi trường huấn luyện mô hình và triển khai mô hình vào ứng dụng di động. Chúng tôi thực nghiệm trên môi trường Google Colab GPU Tesla T4.

5.1. Thiết lập dữ liệu và tổ chức thư mục

YOLOv5 yêu cầu 2 thư mục: 1 thư mục để huấn luyện và 1 thư mục để xác thực. Trong mỗi 2 thư mục đó, cần thêm 2 thư mục nữa là “Images” và “Labels”. Thư mục “Image” chứa các hình ảnh thực tế. Thư mục “Labels” chứa tệp .txt cho mỗi hình ảnh với chú thích của hình ảnh đó. Cấu trúc thư mục được tổ chức như hình 25.



Hình 25. Cấu trúc thư mục dữ liệu

5.2. Huấn luyện mô hình YOLOv5

Nghiên cứu này sử dụng mô hình YOLOv5m. Mô hình được đào tạo trên Colab sử dụng GPU Tesla T4 với các thông số mặc định và được sửa đổi như bảng 3.

Hyperparameter	Cài đặt		Hyperparameter	Cài đặt	
	Mặc định	Đã sửa đổi		Mặc định	Đã sửa đổi
lr0	0.01	0.001	box	0.05	0.05
lrf	0.1	0.1	cls	0.5	0.5
momentum	0.937	0.937	cls_pw	1.0	1.0
Weight_decay	0.0005	0.0005	obj	1.0	1.0

Hyperparameter	Cài đặt		Hyperparameter	Cài đặt	
	Mặc định	Đã sửa đổi		Mặc định	Đã sửa đổi
img	640	416	obj_pw	1.0	1.0
Batch size	32	16	Iou_t	0.20	0.20
epochs	100	200			

Bảng 2. Các Hyperparameters mặc định và được sửa đổi để huấn luyện mô hình YOLOv5

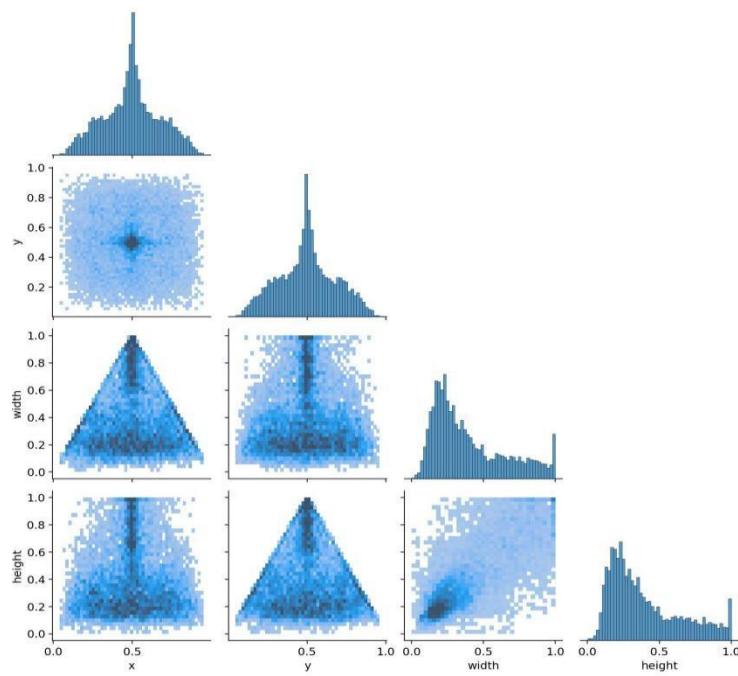
❖ Kết quả train mô hình YOLOv5

STT	Tên đối tượng	Precision	Recall	mAP
1	Bánh Bò Chăm	0.796	0.5	0.692
2	Bánh Bò Thốt Nốt	0.704	0.659	0.72
3	Bánh Cam	0.789	0.884	0.869
4	Bánh Canh Ghe	0.763	0.96	0.969
5	Bánh Còng	0.767	0.75	0.837
6	Bánh Cống	0.867	0.784	0.899
7	Bánh Chuối	0.997	0.933	0.991
8	Bánh Đúc Lá Dứa	0.957	0.643	0.813
9	Bánh Đúc Mặn	0.987	0.909	0.907
10	Bánh Khot	0.89	0.87	0.928
11	Bánh Lá Dừa	0.657	0.295	0.502
12	Bánh Ông Lá Dứa	0.488	0.578	0.335
13	Bánh Pía	0.884	0.847	0.943
14	Bánh Tai Yên	0.849	0.739	0.836
15	Bánh Tầm Bì	0.838	0.69	0.88
16	Bánh Tầm Cay	0.899	1	0.979
17	Bánh Tầm Khoai Mì	0.927	0.733	0.813
18	Bánh Tét Lá Cẩm	0.915	0.877	0.975
19	Bánh Xèo	0.843	0.828	0.883
20	Bún Bò Cay	0.733	1	0.995
21	Bún Cá	0.902	0.837	0.91
22	Bún Kèn Hà Tiên	0.87	0.833	0.959

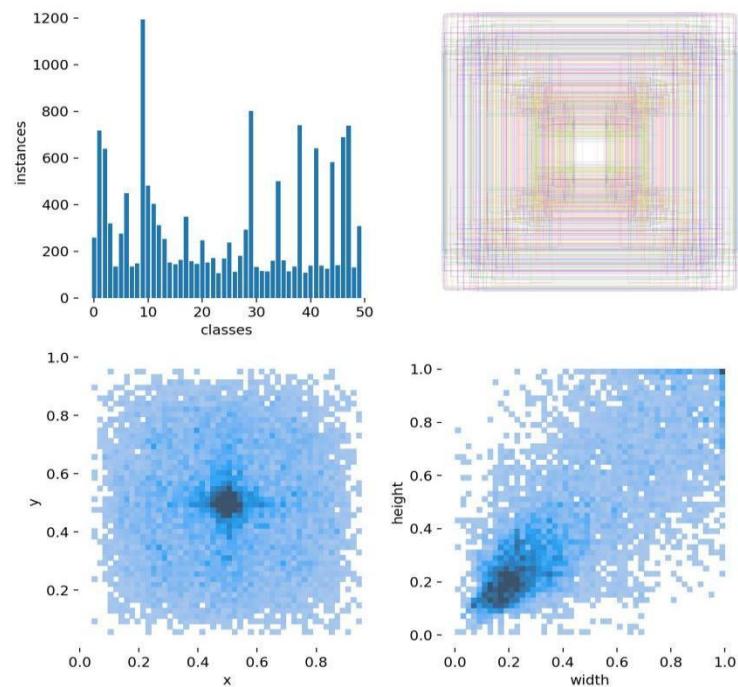
STT	Tên đối tượng	Precision	Recall	mAP
23	Bún Quậy	0.962	0.941	0.954
24	Cá Kèo Nướng Muối Ót	1	0.814	0.829
25	Cá Kho Tộ	0.879	0.912	0.918
26	Cá Lóc Nướng	0.912	0.543	0.86
27	Cá Tai Tượng Chiên Xù	1	0.95	0.995
28	Canh Chua	0.977	0.941	0.982
29	Cơm Cháy Kho Quẹt	0.537	0.375	0.341
30	Cơm Gói Lá Sen	0.906	0.875	0.938
31	Cơm Tấm Long Xuyên	1	0.802	0.931
32	Chè Bưởi	0.801	0.803	0.887
33	Chuối Nép Nướng	0.901	0.814	0.931
34	Dừa Sáp	0.708	1	0.995
35	Đuông Dừa	0.65	0.495	0.551
36	Gà Đốt Ô Thum	0.877	0.933	0.961
37	Gỏi Ba Khía	0.957	1	0.995
38	Gỏi Cá Trích	0.754	0.692	0.825
39	Gỏi Cuốn	0.841	0.442	0.674
40	Gỏi Sầu Đâu	0.753	0.5	0.676
41	Hủ Tiếu Mỹ Tho	0.761	0.909	0.819
42	Kẹo Dừa	0.786	0.821	0.848
43	Lẩu Gà Chanh Ót	0.92	0.891	0.917
44	Lẩu Mắm	0.923	0.998	0.95
45	Nem Lai Vung	0.765	0.481	0.686
46	Ốc Len Xào Dừa	0.973	1	0.995
47	Óc Nướng Tiêu Xanh	0.708	0.941	0.934
48	Thốt Nốt	0.586	0.729	0.664
49	Vịt Nấu Chao	0.837	0.833	0.958
50	Xôi Phòng	0.789	0.692	0.804

Bảng 3. Kết quả huấn luyện mô hình YOLOv5

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ

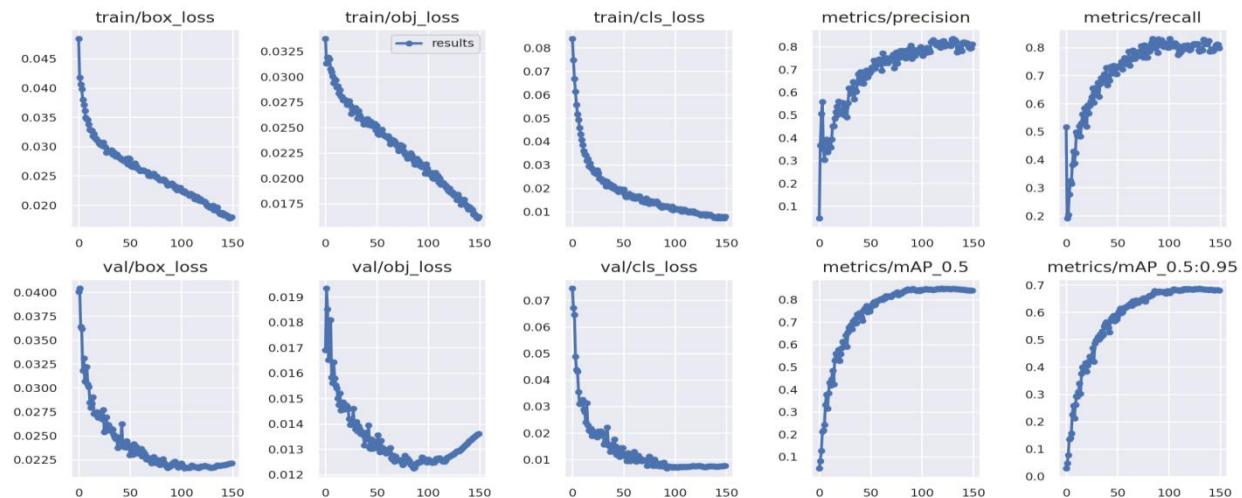


Hình 26. Lables correlogram

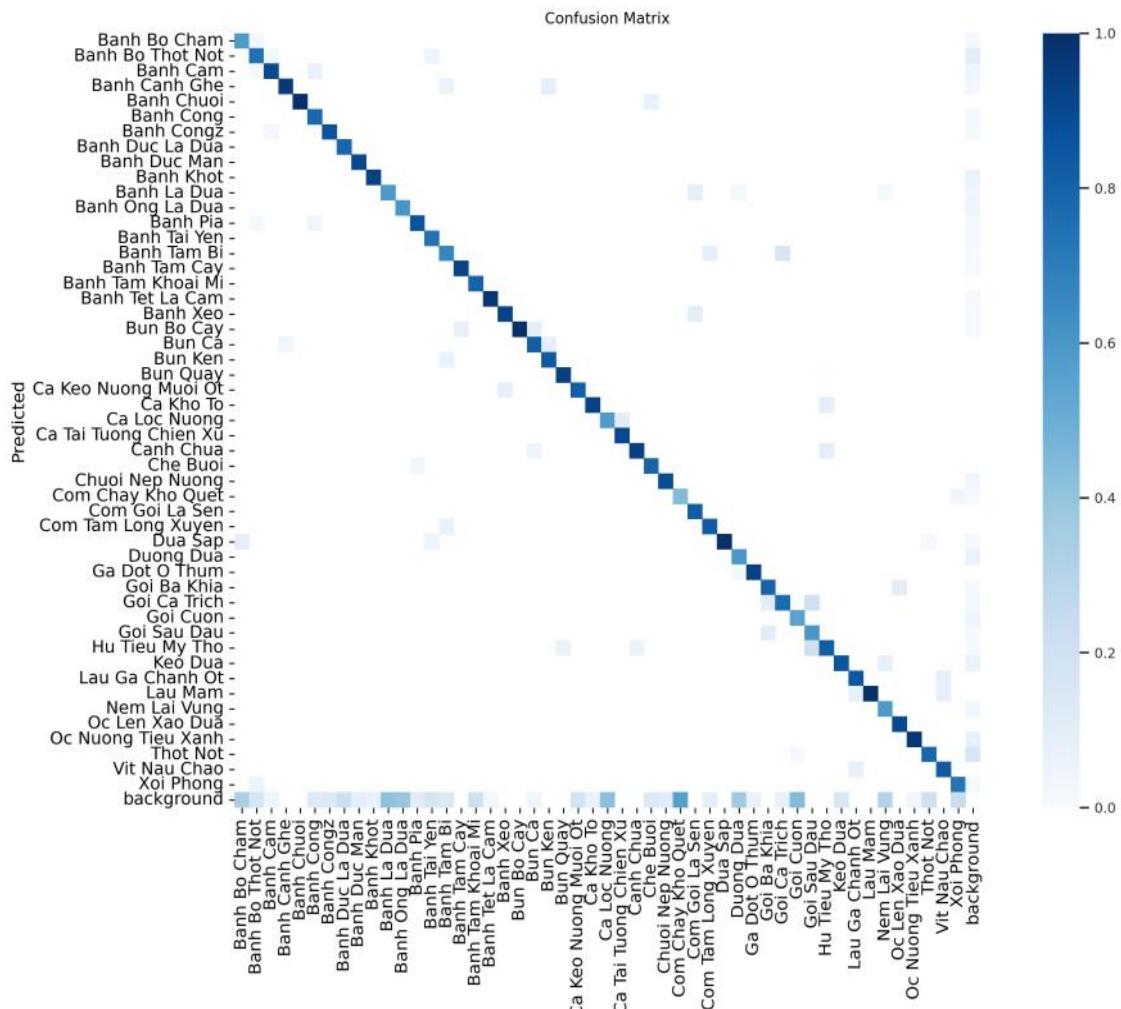


Hình 27. Lables

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ



Hình 28. Kết quả train



Hình 29. Ma trận nhầm lẫn

❖ Kết quả 3 mô hình YOLOv5, YOLOv7 và ViT

STT	Tên đối tượng	mAP			
		MoNV3	YOLOv5	YOLOv7	ViT
1	Bánh Bò Chăm	0.75	0.692	0.860	0.9
2	Bánh Bò Thốt Nốt	0.874	0.72	0.737	0.9
3	Bánh Cam	0.75	0.869	0.879	0.8
4	Bánh Canh Ghẹ	0.759	0.969	0.975	1
5	Bánh Còng	0.91	0.837	0.881	0.9
6	Bánh Công	0.750	0.899	0.893	0.75
7	Bánh Chuối	0.820	0.991	0.898	1
8	Bánh Đúc Lá Dứa	0.800	0.813	0.916	1
9	Bánh Đúc Mặn	0.727	0.907	0.902	1
10	Bánh Khot	0.772	0.928	0.931	0.9
11	Bánh Lá Dừa	0.778	0.502	0.500	1
12	Bánh Ông Lá Dứa	0.750	0.335	0.413	1
13	Bánh Pía	1.000	0.943	0.967	1
14	Bánh Tai Yên	0.880	0.836	0.769	1
15	Bánh Tầm Bì	0.690	0.88	0.932	1
16	Bánh Tầm Cay	0.500	0.979	0.977	0.88
17	Bánh Tầm Khoai Mì	0.750	0.813	0.878	0.87
18	Bánh Tết Lá Cảm	0.666	0.975	0.984	1
19	Bánh Xèo	0.909	0.883	0.849	0.9
20	Bún Bò Cay	0.560	0.995	0.995	1
21	Bún Cá	0.769	0.91	0.924	1
22	Bún Kèn Hà Tiên	0.800	0.959	0.924	1
23	Bún Quậy	0.857	0.954	0.993	0.88
24	Cá Kèo Nướng Muối Ót	0.706	0.829	0.956	1
25	Cá Kho Tộ	0.800	0.918	0.988	0.9
26	Cá Lóc Nướng	0.692	0.86	0.874	0.9
27	Cá Tai Tượng Chiên Xù	0.842	0.995	0.956	0.77
28	Canh Chua	0.516	0.982	0.927	1
29	Cơm Cháy Kho Quẹt	0.750	0.341	0.384	1
30	Cơm Gói Lá Sen	0.693	0.938	0.871	0.9

STT	Tên đồi tượng	mAP			
		MoNV3	YOLOv5	YOLOv7	ViT
31	Cơm Tấm Long Xuyên	0.770	0.931	0.995	0.9
32	Chè Bưởi	0.810	0.887	0.902	1
33	Chuối Nép Nướng	0.900	0.931	0.900	0.9
34	Dừa Sáp	0.846	0.995	0.993	1
35	Đuông Dừa	0.769	0.551	0.544	1
36	Gà Đốt Ô Thum	0.786	0.961	0.961	1
37	Gỏi Ba Khía	0.667	0.995	0.995	1
38	Gỏi Cá Trích	0.527	0.825	0.969	1
39	Gỏi Cuốn	0.936	0.674	0.791	1
40	Gỏi Sầu Đâu	0.544	0.676	0.782	0.8
41	Hủ Tiếu Mỹ Tho	0.667	0.819	0.902	1
42	Kéo Dừa	0.917	0.848	0.892	0.9
43	Lẩu Gà Chanh Ót	0.800	0.917	0.792	1
44	Lẩu Mắm	0.706	0.95	0.984	1
45	Nem Lai Vung	0.727	0.686	0.662	0.9
46	Óc Len Xào Dừa	0.827	0.995	0.912	0.9
47	Óc Nướng Tiêu Xanh	0.897	0.934	0.923	1
48	Trái Thốt Nốt	0.941	0.664	0.684	1
49	Vịt Nấu Chao	0.800	0.958	0.985	0.9
50	Xôi Phòng	0.957	0.804	0.795	1

Bảng 4. mAP của các mô hình nhận dạng món ăn MobileNet-V3(MoNV3), YOLOv5, YOLOv7, và Vision Transformer(ViT)

Sau khi, chúng tôi tinh chỉnh và huấn luyện các mô hình nhận dạng món ăn. Kết quả mAP của các mô hình vision Transformer, MobileNetV3, YOLOv5, và YOLOv7 lần lượt là 0,931; 0,772; 0,849 và 0,870.

5.3. Triển khai mô hình bằng TensorFlow Lite

Tệp “best.pt” được xuất thành file .tflite để triển khai trên một ứng dụng di động nhằm mục đích thử nghiệm. Ứng dụng di động dựa trên TensorFlow được xây dựng và chạy trên môi trường phát triển Android Studio.

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ



Hình 30. Giao diện trang chủ của ứng dụng với tiếng Việt



Hình 31. Giao diện trang chủ của ứng dụng với tiếng Anh

Ứng dụng cho phép người dùng sử dụng 2 ngôn ngữ: tiếng Việt và tiếng Anh. Từ giao diện chúng ta có thể chọn nút “Nhận diện trực tiếp” (“Live Detection”) để nhận dạng ảnh real-time như hình 32, hình 33.



Hình 32. Giao diện khi chọn chức năng “Nhận dạng trực tiếp”



Hình 33. Giao diện khi chọn chức năng “Live Detection”

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

Hay có thể từ giao diện chọn chức năng “Nhận dạng qua ảnh” (“Gallery Detection”). Sẽ có 2 lựa chọn “Tải ảnh” (“Upload Image”) từ thư mục ảnh của người dùng, hoặc “Camera” chụp ảnh trực tiếp từ máy ảnh.

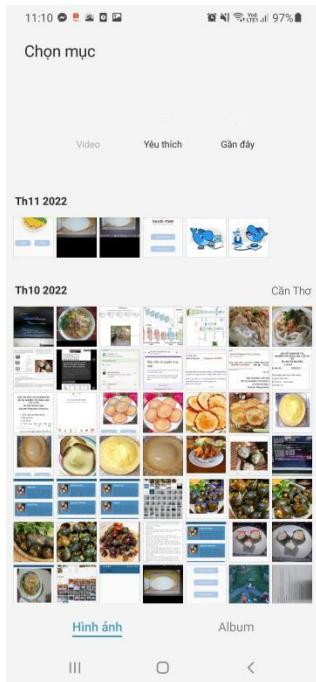


Hình 34. Giao diện khi chọn chức năng “Nhận dạng qua ảnh”



Hình 35. Giao diện khi chọn chức năng “Gallery Detection”

Nếu chọn chức năng “Tải ảnh” (“Upload Image”) hệ thống sẽ mở thư mục ảnh trên điện thoại cho người dùng lựa chọn ảnh tải lên.

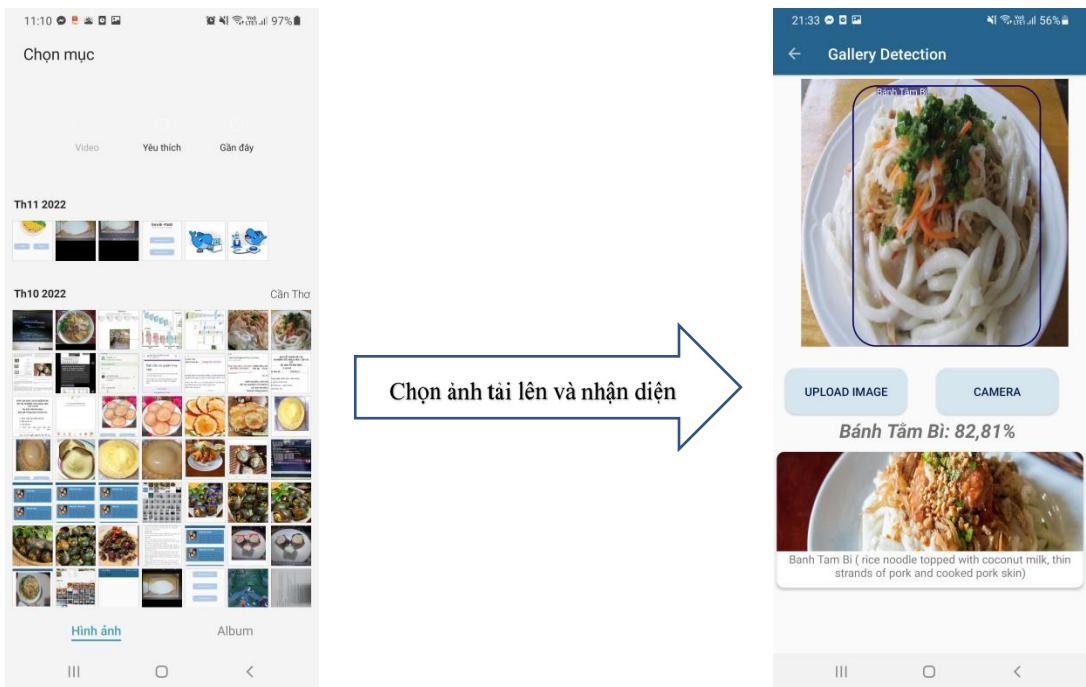


Chọn ảnh tải lên và nhận diện



Hình 36. Giao diện ứng dụng khi chọn chức năng “Tải ảnh”

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ



Hình 37. Giao diện ứng dụng khi chọn chức năng “Upload Image”

Nếu chọn chức năng “Camera” hệ thống sẽ mở máy ảnh của điện thoại cho người dùng chụp ảnh.



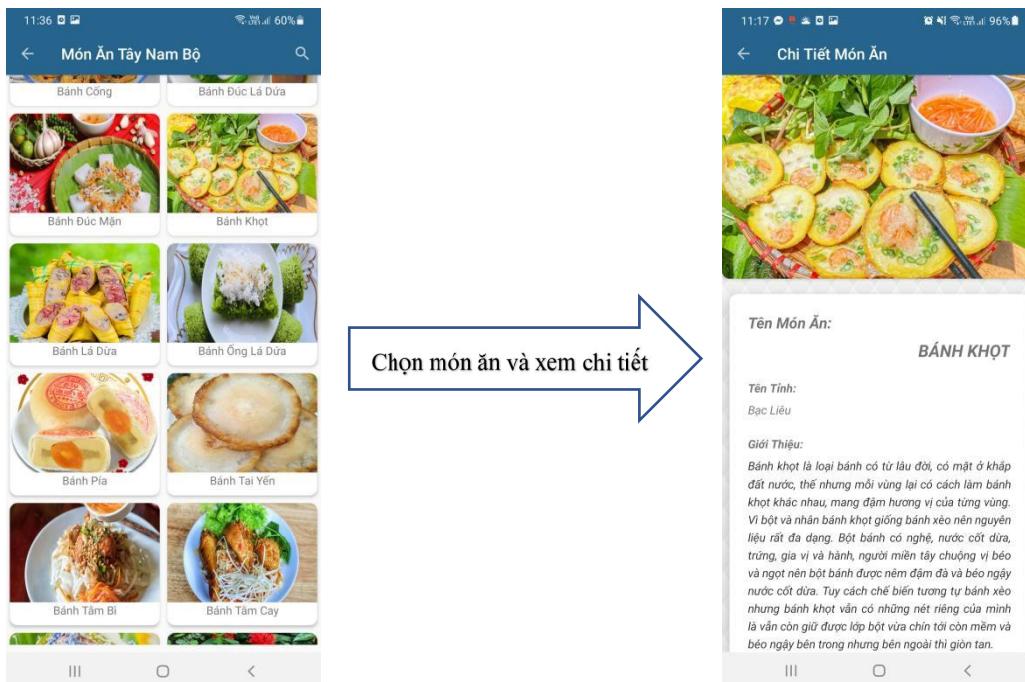
Hình 38. Giao diện ứng dụng khi chọn chức năng “Camera” với tiếng Việt

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ



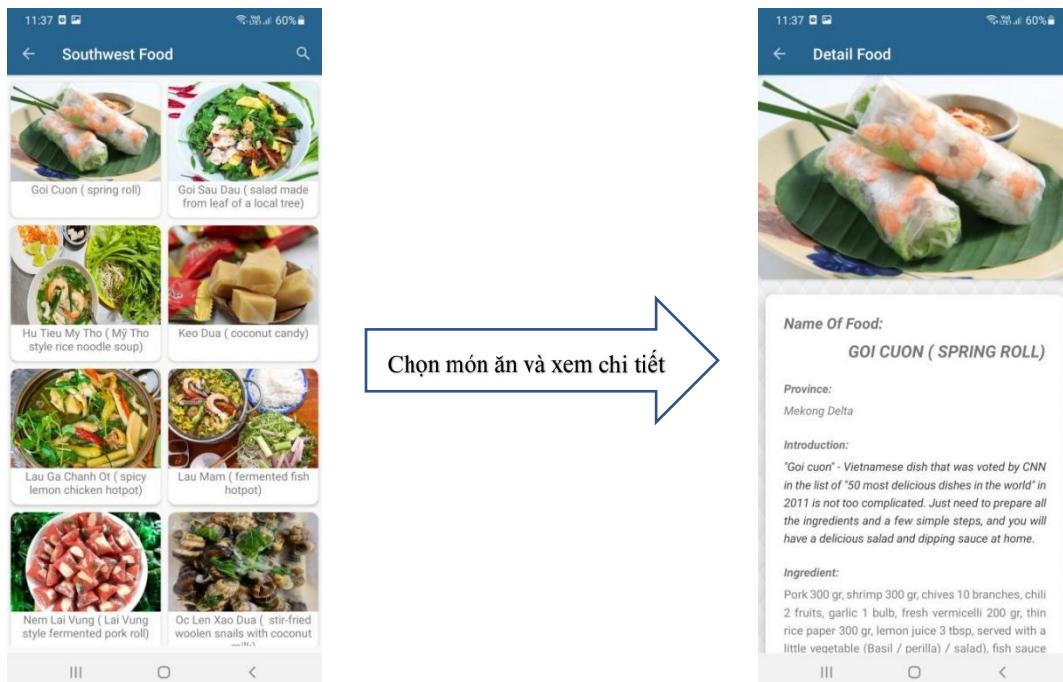
Hình 39. Giao diện ứng dụng khi chọn chức năng “Camera” với tiếng Anh

Hoặc chọn chức năng “Thông tin món ăn” (“Detail Food”) để tham khảo hoặc tìm hiểu thông tin các món ăn khác. Hay tìm kiếm một món ăn nào đó.

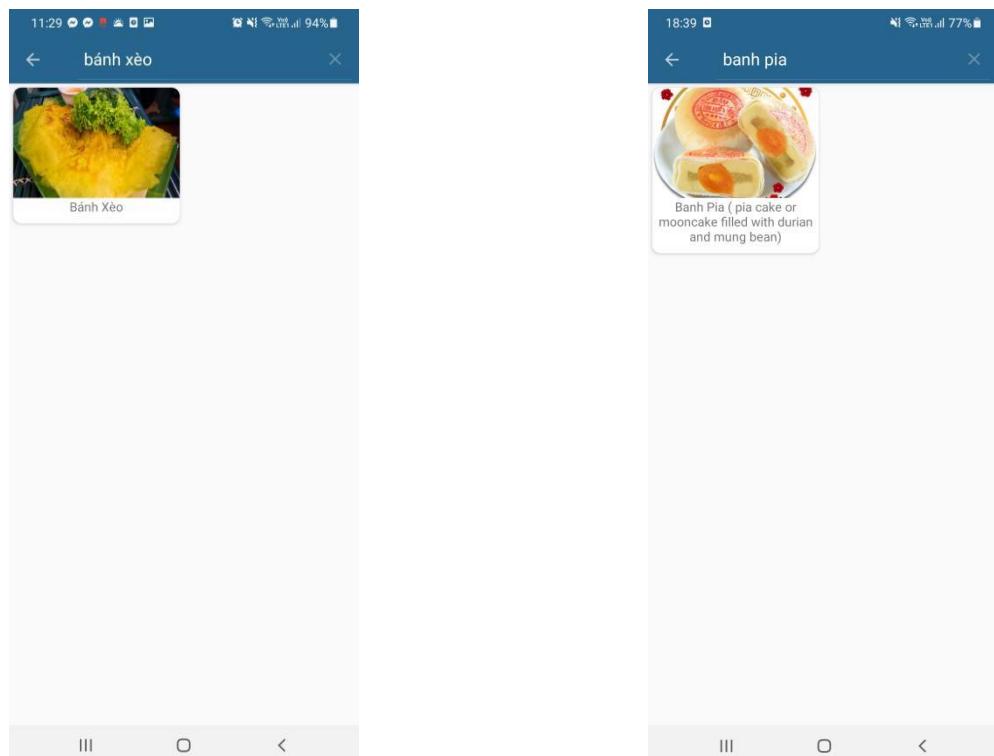


Hình 40. Giao diện ứng dụng khi chọn chức năng “Thông tin món ăn”

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ



Hình 41. Giao diện ứng dụng khi chọn chức năng “Detail Food”



Hình 42. Giao diện ứng dụng chức năng tìm kiếm với tiếng Việt

Hình 43. Giao diện ứng dụng chức năng tìm kiếm với tiếng Anh

CHƯƠNG 6. KẾT LUẬN

6.1. Kết quả đạt được

Về cơ bản, đề tài đã thu thập được tập dữ liệu các món ăn đặc trưng Tây Nam Bộ. Biết cách dùng Roboflow gắn nhãn cũng như áp dụng các kỹ thuật tăng cường cho dữ liệu. Áp dụng được các mô hình nhận dạng đối tượng vào quá trình huấn luyện mô hình. Tích hợp thành công mô hình nhận dạng đối tượng vào ứng dụng di động, cho phép nhận dạng, tìm kiếm các món ăn trong tập dữ liệu đã thu thập. Tuy nhiên, còn một số hạn chế: số lượng món ăn trong tập dữ liệu còn bị giới hạn, độ chính xác chưa cao vì các món ăn khá đa dạng về nguyên liệu, cách trình bày.

6.2. Hướng phát triển

Trong tương lai, nhóm sẽ cải thiện khả năng nhận dạng của mô hình cả về số lượng món ăn và độ chính xác. Thiết kế giao diện ứng dụng chuyên nghiệp và nhiều chức năng hơn tạo cảm giác thân thiện với người dùng. Và cho phép người dùng cung cấp dataset cho ứng dụng.

TÀI LIỆU THAM KHẢO

- [1] Hoashi, Hajime, Taichi Joutou, and Keiji Yanai, "Image recognition of 85 food categories by feature fusion," in *IEEE International Symposium on Multimedia*, 2010.
- [2] Kagaya, Hokuto, Kiyoharu Aizawa, and Makoto Ogawa, "Food detection and recognition using convolutional neural network," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014.
- [3] Y. Lu, "Food image recognition by using convolutional neural networks (CNNs)," *arXiv preprint arXiv:1612.00983*, 2016.
- [4] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *nature*, vol. 521, p. 436–444, 2015.
- [5] A. A. Jeny, M. S. Junayed, I. Ahmed, M. T. Habib and M. R. Rahman, "FoNet-Local food recognition using deep residual neural networks," in *2019 International Conference on Information Technology (ICIT)*, 2019.
- [6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.
- [7] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, "Mobilennets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [8] Ş. Aktı, M. Qaraqe and H. K. Ekenel, "A Mobile Food Recognition System for Dietary Assessment," *arXiv preprint arXiv:2204.09432*, 2022.
- [9] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference*

on computer vision and pattern recognition, 2018.

- [10] Z. Shen, A. Shehzad, S. Chen, H. Sun and J. Liu, "Machine learning based approach on food recognition and nutrition estimation," *Procedia Computer Science*, vol. 174, p. 448–453, 2020.
- [11] Z. Zahisham, C. P. Lee and K. M. Lim, "Food recognition with Resnet-50," in *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 2020.
- [12] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [13] W. Min, L. Liu, Z. Wang, Z. Luo, X. Wei, X. Wei and S. Jiang, "Isia food-500: A dataset for large-scale food recognition via stacked global-local attention network," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.
- [14] W. Min, Z. Wang, Y. Liu, M. Luo, L. Kang, X. Wei, X. Wei and S. Jiang, "Large scale visual food recognition," *arXiv preprint arXiv:2103.16107*, 2021.
- [15] L. Bossard, M. Guillaumin and L. V. Gool, "Food-101—Mining discriminative components with Random Forests," in *European conference on computer vision*, 2014.
- [16] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton., "Imagenet classification with deep convolutional neural networks," *Communications of the ACM* 60, vol. 60, no. 6, pp. 84-90, 2017.
- [17] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun., "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015.
- [18] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards

- real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [19] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [20] H. T. Ung, T. X. Dang, P. V. Thai, T. T. Nguyen and B. T. Nguyen, "Vietnamese food recognition system using convolutional neural networks based features," in *International Conference on Computational Collective Intelligence*, 2020.
- [21] T. T. Nguyen, T. Q. Nguyen, D. Vo, V. Nguyen, N. Ho, N. D. Vo, K. Van Nguyen and K. Nguyen, "VinaFood21: A Novel Dataset for Evaluating Vietnamese Food Recognition," in *2021 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2021.
- [22] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019.
- [23] T.-H. Do, H.-Q. Dang, H.-N. Nguyen, P.-P. Pham, D.-T. Nguyen and others, "30VNFood: A Dataset for Vietnamese Foods Recognition," in *2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2021.
- [24] C.-Y. Wang, A. Bochkovskiy and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696*, 2022.
- [25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly and others, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

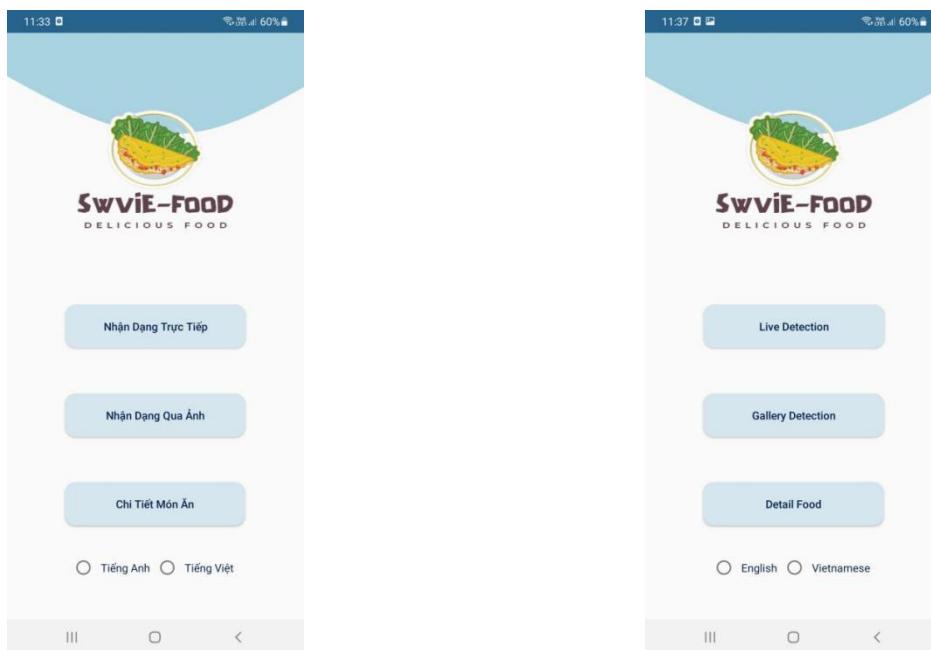
- [26] Howard, Andrew, et al. "Searching for mobilenetv3." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.
- [27] Sultonov, Furkat, et al. "Mixer U-Net: An Improved Automatic Road Extraction from UAV Imagery." *Applied Sciences* 12.4 (2022): 1953.
- [28] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [29] Xu, Renjie, Haifeng Lin, Kangjie Lu, Lin Cao, and Yunfei Liu, "Forest Fire Detection System Based on Ensemble Learning," *Forests*, vol. 12, no. 2, p. 217, 2021.

PHỤ LỤC

SẢN PHẨM

Sản phẩm là ứng dụng di động nhận diện ẩm thực miền Tây Nam Bộ.

Ứng dụng di động dựa trên TensorFlow được xây dựng và chạy trên môi trường phát triển Android Studio.



Giao diện trang chủ của ứng dụng với tiếng Việt

Giao diện trang chủ của ứng dụng với tiếng Anh

Ứng dụng cho phép người dùng sử dụng 2 ngôn ngữ: tiếng Việt và tiếng Anh. Từ giao diện chúng ta có thể chọn nút “Nhận diện trực tiếp” (“Live Detection”) để nhận dạng ảnh real-time.

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ



Giao diện khi chọn chức năng “Nhận dạng trực tiếp”



Giao diện khi chọn chức năng “Live Detection”

Hay có thể từ giao diện chọn chức năng “Nhận dạng qua ảnh” (“Gallery Detection”). Sẽ có 2 lựa chọn “Tải ảnh” (“Upload Image”) từ thư mục ảnh của người dùng, hoặc “Camera” chụp ảnh trực tiếp từ máy ảnh.



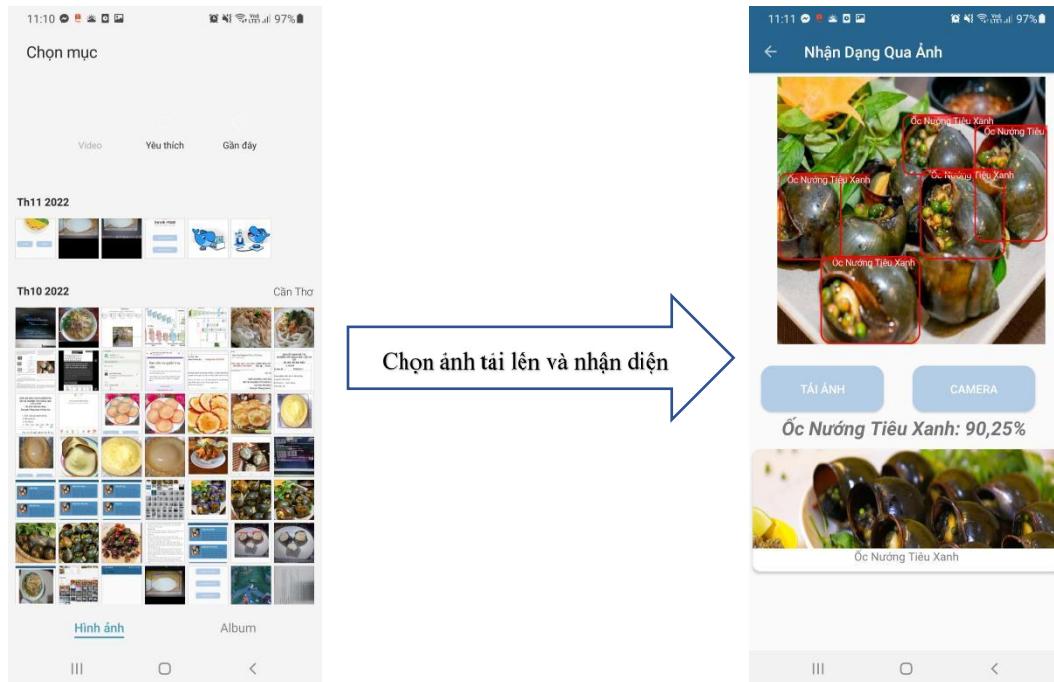
Giao diện khi chọn chức năng “Nhận dạng qua ảnh”



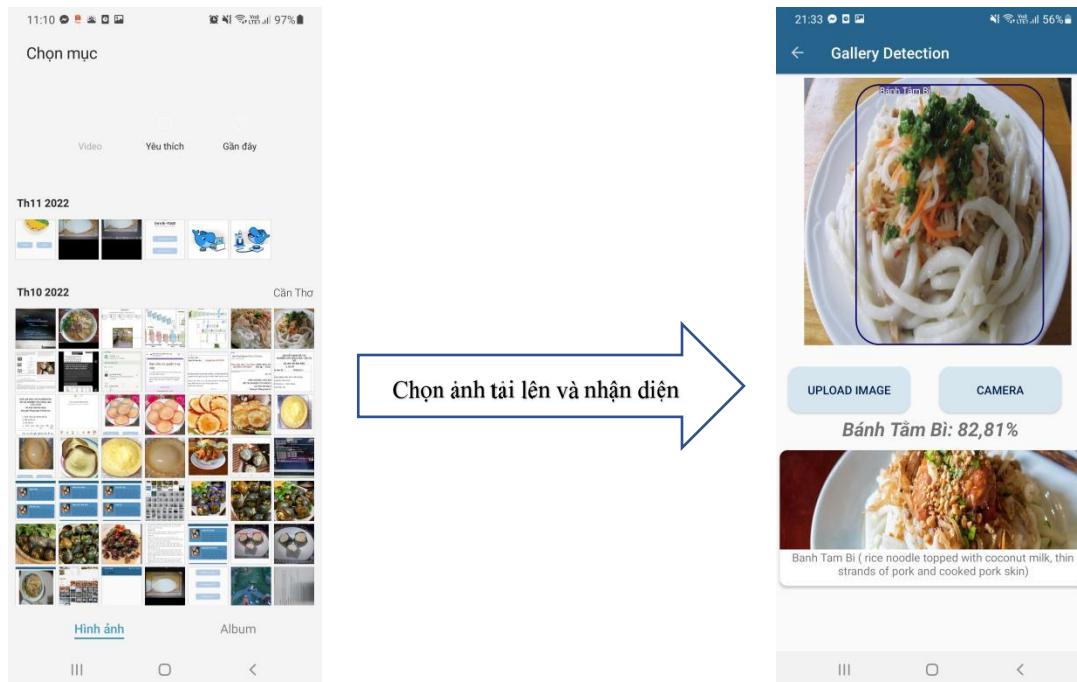
Giao diện khi chọn chức năng “Gallery Detection”

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

Nếu chọn chức năng “Tải ảnh” (“Upload Image”) hệ thống sẽ mở thư mục ảnh trên điện thoại cho người dùng lựa chọn ảnh tải lên.



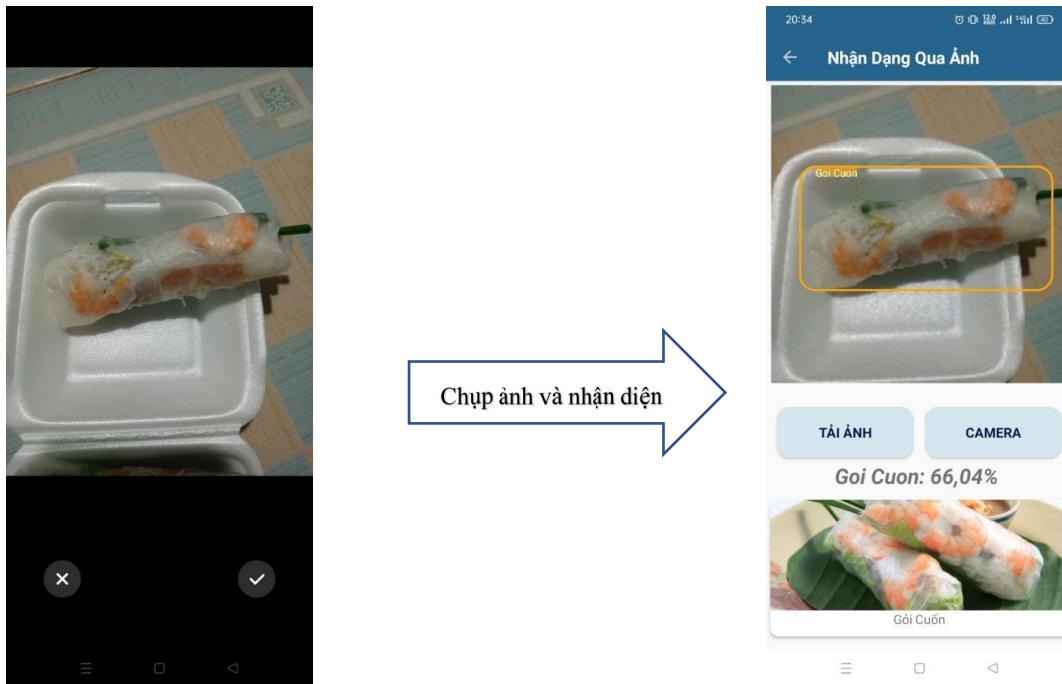
Giao diện ứng dụng khi chọn chức năng “Tải ảnh”



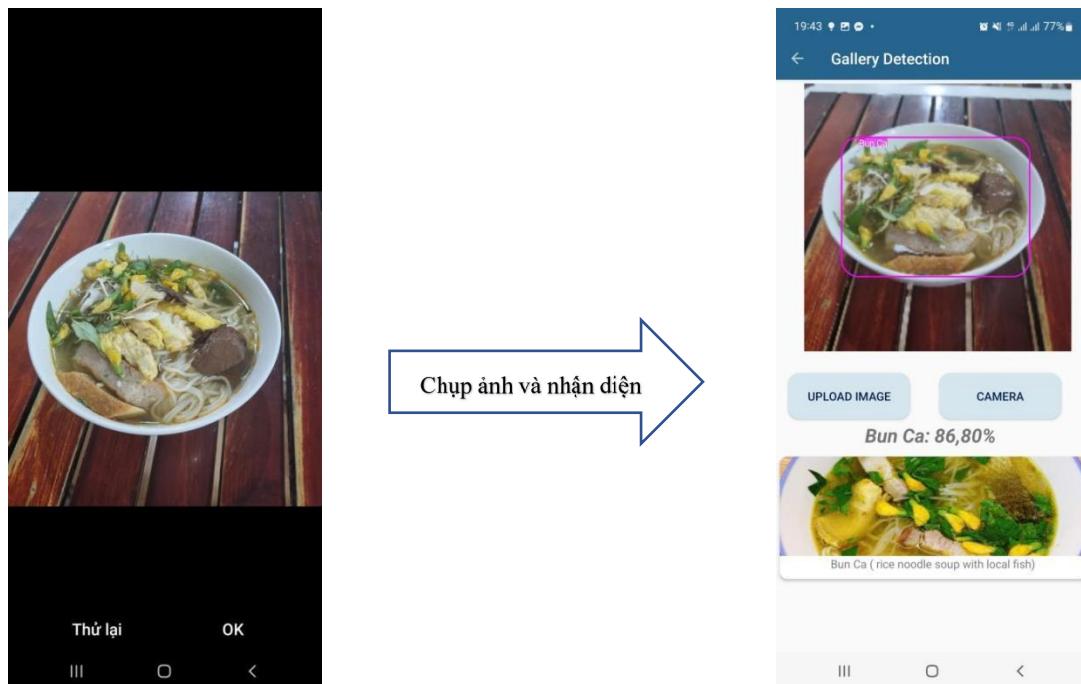
Giao diện ứng dụng khi chọn chức năng “Upload Image”

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

Nếu chọn chức năng “Camera” hệ thống sẽ mở máy ảnh của điện thoại cho người dùng chụp ảnh.



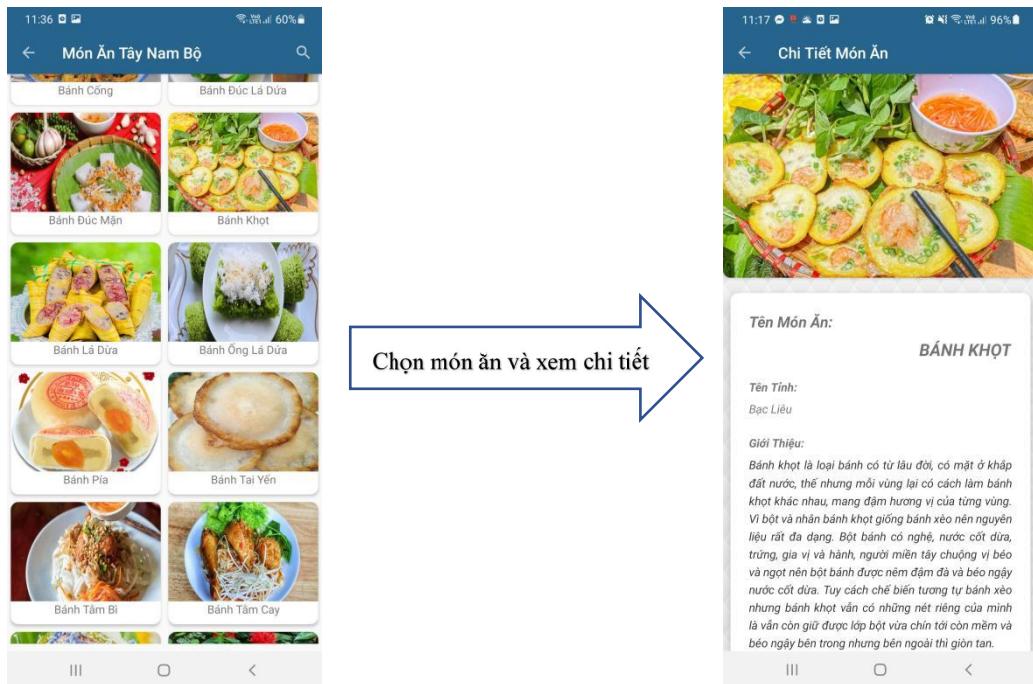
Giao diện ứng dụng khi chọn chức năng “Camera” với tiếng Việt



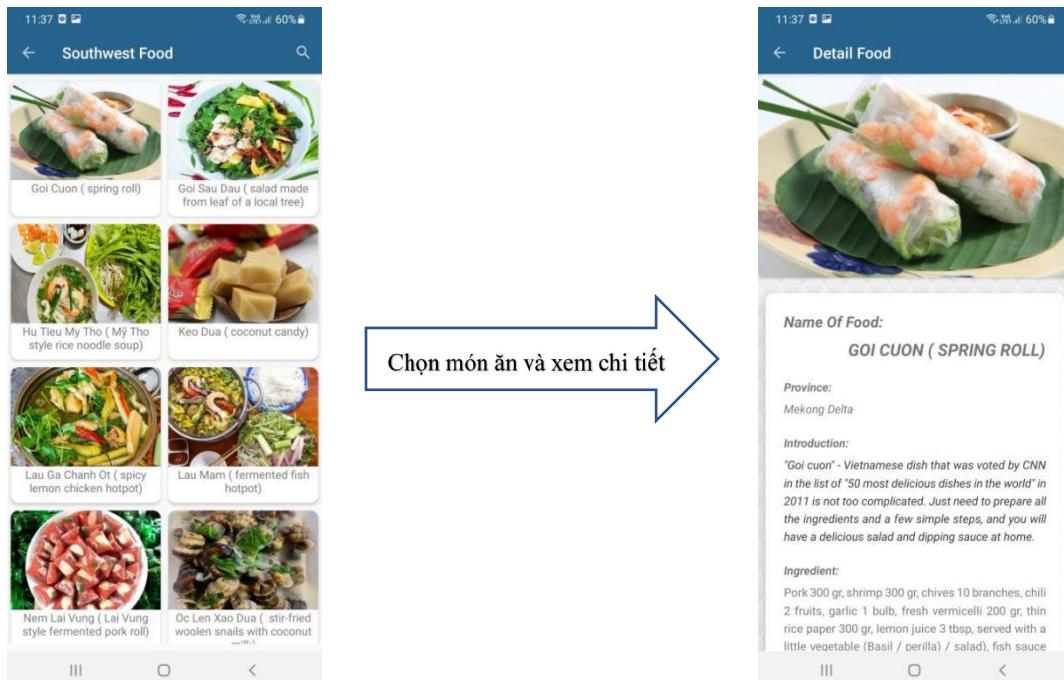
Giao diện ứng dụng khi chọn chức năng “Camera” với tiếng Anh

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

Hoặc chọn chức năng “Thông tin món ăn” (“Detail Food”) để tham khảo hoặc tìm hiểu thông tin các món ăn khác. Hay tìm kiếm một món ăn nào đó.

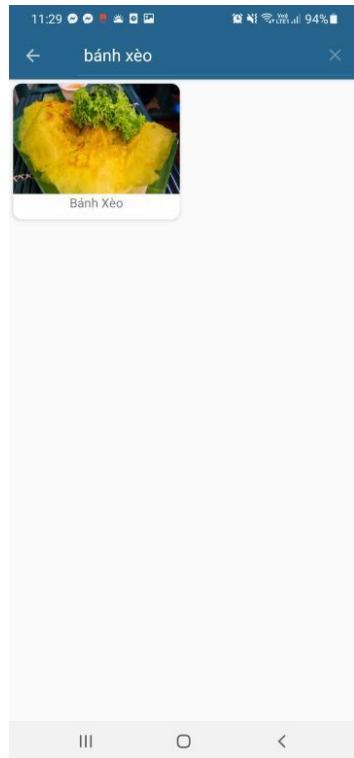


Giao diện ứng dụng khi chọn chức năng “Thông tin món ăn”

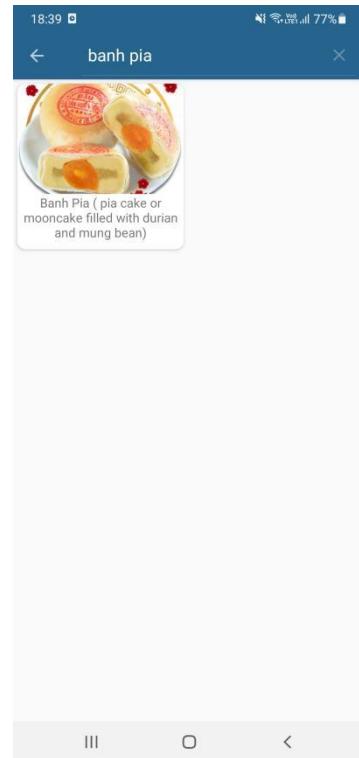


Giao diện ứng dụng khi chọn chức năng “Detail Food”

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ



*Giao diện ứng dụng chức năng tìm kiếm với
tiếng Việt*



*Giao diện ứng dụng chức năng tìm kiếm với
tiếng Anh*

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

<p>1. TÊN ĐỀ TÀI</p> <p>Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ</p>	<p>2. MÃ SỐ</p> <p>THS2022-17</p>																				
<p>Lĩnh vực ưu tiên</p> <ul style="list-style-type: none"> <input type="checkbox"/> Lĩnh vực 1. Ứng dụng công nghệ cao trong nông nghiệp, thủy sản và môi trường <input type="checkbox"/> Lĩnh vực 2. Quản lý và sử dụng bền vững tài nguyên thiên nhiên <input checked="" type="checkbox"/> Lĩnh vực 3. Kỹ thuật công nghệ và công nghệ thông tin – truyền thông <input type="checkbox"/> Lĩnh vực 4. Khoa học Giáo dục, Luật và Xã hội Nhân văn <input type="checkbox"/> Lĩnh vực 5. Phát triển kinh tế, thị trường <input type="checkbox"/> Không thuộc 05 Lĩnh vực ưu tiên. 																					
<p>3. LĨNH VỰC NGHIÊN CỨU</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30%;">Khoa học Tự nhiên</td> <td style="width: 10%;"><input type="checkbox"/></td> <td style="width: 30%;">Khoa học Kỹ thuật và Công nghệ</td> <td style="width: 10%;"><input checked="" type="checkbox"/></td> </tr> <tr> <td>Khoa học Y, dược</td> <td><input type="checkbox"/></td> <td>Khoa học Nông nghiệp</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Khoa học Xã hội</td> <td><input type="checkbox"/></td> <td>Khoa học Nhân văn</td> <td><input type="checkbox"/></td> </tr> </table>	Khoa học Tự nhiên	<input type="checkbox"/>	Khoa học Kỹ thuật và Công nghệ	<input checked="" type="checkbox"/>	Khoa học Y, dược	<input type="checkbox"/>	Khoa học Nông nghiệp	<input type="checkbox"/>	Khoa học Xã hội	<input type="checkbox"/>	Khoa học Nhân văn	<input type="checkbox"/>	<p>4. LOẠI HÌNH NGHIÊN CỨU</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30%;">Cơ bản</td> <td style="width: 10%;"><input type="checkbox"/></td> <td style="width: 30%;">Ứng dụng</td> <td style="width: 10%;"><input type="checkbox"/></td> </tr> <tr> <td>Triển khai</td> <td><input type="checkbox"/></td> <td><input checked="" type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </table>	Cơ bản	<input type="checkbox"/>	Ứng dụng	<input type="checkbox"/>	Triển khai	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Khoa học Tự nhiên	<input type="checkbox"/>	Khoa học Kỹ thuật và Công nghệ	<input checked="" type="checkbox"/>																		
Khoa học Y, dược	<input type="checkbox"/>	Khoa học Nông nghiệp	<input type="checkbox"/>																		
Khoa học Xã hội	<input type="checkbox"/>	Khoa học Nhân văn	<input type="checkbox"/>																		
Cơ bản	<input type="checkbox"/>	Ứng dụng	<input type="checkbox"/>																		
Triển khai	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>																		
<p>5. THỜI GIAN THỰC HIỆN</p> <p>06 tháng</p> <p>Từ tháng 06 năm 2022 đến tháng 11 năm 2022</p>																					

Đề tài: Xây dựng ứng dụng di động giới thiệu âm thực miền Tây Nam Bộ

6. ĐƠN VỊ CỦA CHỦ NHIỆM ĐỀ TÀI

Tên đơn vị: Khoa CNTT&TT Trường Đại học Cần Thơ

Điện thoại: 02923734713

E-mail: office@cit.ctu.edu.vn

Địa chỉ: Khu 2, đường 3/2, Phường Xuân Khánh, Quận Ninh Kiều, TP. Cần Thơ, Việt Nam

Họ và tên thủ trưởng đơn vị: TS. Nguyễn Hữu Hòa

7. CHỦ NHIỆM ĐỀ TÀI

Họ và tên: Nguyễn Thị Mỹ Khanh MSSV: B1910657

Ngày tháng năm sinh: 04-05-2001 Lớp: Công nghệ thông tin-Chất lượng cao 2

Điện thoại di động: 0976038762 Khóa: 45

E-mail: khanhb1910657@student.ctu.edu.vn

8. NHỮNG THÀNH VIÊN THAM GIA NGHIÊN CỨU ĐỀ TÀI

TT	Họ và tên	MSSV, Lớp, Khóa	Nội dung nghiên cứu cụ thể được giao	Chữ ký
1	Nguyễn Thị Mỹ Khanh (Chủ nhiệm đề tài)	MSSV: B1910657 Lớp: Công nghệ thông tin-Chất lượng cao 2 Khóa: 45	- Nghiên cứu lý thuyết về mô hình YoloV5 và các nghiên cứu liên quan đến ứng dụng di động. - Thu nhập và xây dựng tập dữ liệu mô hình. - Viết báo cáo tổng kết.	
2	Nguyễn Duy Khang (Thành viên chính)	MSSV: B1910654 Lớp: Công nghệ thông tin-Chất lượng cao 2 Khóa: 45	- Nghiên cứu lý thuyết về mô hình YoloV5 và các nghiên cứu liên quan đến ứng dụng di động. - Thu nhập và xây dựng tập dữ liệu mô hình. - Viết báo cáo tổng kết.	
3	Nguyễn Hiếu Nghĩa (Thành viên chính)	MSSV: B1910672 Lớp: Công nghệ thông tin-Chất lượng cao 2 Khóa: 45	- Thu nhập và xây dựng tập dữ liệu mô hình. - Lập trình, xây dựng, thiết kế ứng dụng di động. - Viết báo cáo tổng kết.	

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

4	Tống Phước Lộc (Thành viên chính)	MSSV: B1910664 Lớp: Công nghệ thông tin-Chất lượng cao 2 Khóa: 45	- Thu nhập và xây dựng tập dữ liệu mô hình. - Lập trình, xây dựng, thiết kế ứng dụng di động. - Viết báo cáo tổng kết.	<i>Lộc</i>
5	Lê Hải Yên (Thành viên chính)	MSSV: B1910731 Lớp: Công nghệ thông tin-Chất lượng cao 2 Khóa: 45	- Thu nhập và xây dựng tập dữ liệu mô hình. - Viết báo cáo tổng kết. - Lập trình, xây dựng, thiết kế ứng dụng di động.	<i>Yen</i>

Cán bộ hướng dẫn sinh viên thực hiện đề tài

Họ và tên, MSCB	Đơn vị công tác và lĩnh vực chuyên môn	Nhiệm vụ	Chữ ký
TS. Lâm Nhựt Khang MSCB: 1943	Bộ môn Công nghệ thông tin, Khoa CNTT&TT Lĩnh vực chuyên môn: Khoa học máy tính	Hướng dẫn nội dung khoa học và Hướng dẫn lập dự toán kinh phí đề tài	<i>LN</i>

9. ĐƠN VỊ PHỐI HỢP CHÍNH

Tên đơn vị trong và ngoài nước	Nội dung phối hợp nghiên cứu	Họ và tên người đại diện đơn vị
Không	Không	Không

10. TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU THUỘC LĨNH VỰC CỦA ĐỀ TÀI Ở TRONG VÀ NGOÀI NƯỚC

10.1. Trong nước

Hiện nay, lĩnh vực công nghệ thông tin nước ta đã và đang có những bước phát triển vượt bậc trên mọi lĩnh vực. Các trang web giới thiệu ẩm thực trong nước được giới thiệu một cách chung chung, kết hợp cung cấp các thông tin khác [1], [2], [3]. Chúng tôi chưa tìm thấy ứng dụng di động có tích hợp nhận dạng món ăn đặc trưng vùng Tây Nam Bộ. Bài toán nhận diện và phân loại món ăn nói riêng, nhận diện và phân loại đối tượng nói chung có thể được giải quyết bằng mô hình học sâu, như mô hình Yolo. Các nghiên cứu trong nước đã sử dụng mô hình YoloV4 để nhận diện và phân loại phương tiện giao thông [4], mô hình YoloV5 cũng được sử dụng để giúp nhận dạng tự động các giai đoạn sinh trưởng của cây dưa lưới trong quá trình sinh trưởng trong nhà màng [5].

10.2. Ngoài nước

Theo kiến thức của chúng tôi, không có quá nhiều nghiên cứu liên quan đến bài toán phát hiện và nhận diện món ăn nói chung và món ăn Việt Nam nói riêng. Hoashi và các cộng sự [6] đã sử dụng máy học để xây dựng website cho phép nhận diện 85 món Nhật bản với độ chính xác 62.5%, Kagaya và các cộng sự [7] sử dụng mô hình CNN để phát hiện và nhận diện. Hiện tại, với sự phát triển mạnh mẽ của các mô hình học sâu, các mô hình khác nhau đã được sử dụng để phát hiện và nhận diện đối tượng có thể kể đến như sử dụng mô hình Convolutional Neural Network - CNN [8], Spatial Pyramid Pooling Networks - SPPNet [9], Faster RCNN [10], và Yolo [11]. Mô hình Yolo đạt kết quả vượt trội so với các mô hình học sâu khác. Do đó, trong nghiên cứu này, chúng tôi sẽ nghiên cứu sử dụng mô hình YoloV5 để phát hiện và nhận diện một số món ăn đặc trưng của miền Tây Nam Bộ.

TÀI LIỆU THAM KHẢO

- [1] https://hoangviettravel.vn/dac-san-mien-tay/#2_Mon_an_dac_san_mien_Tay_Nam_Bo
- [2] <https://vinpearl.com/vi/ngat-ngay-voi-top-28-dac-san-mien-tay-ngon-kho-cuong>
- [3] <https://pasgo.vn/blog/top-15-dac-san-mien-tay-nam-bo-lam-quy-duoc-chon-mua-nhieu-nhat-4000>
- [4] Cường, TS Nguyễn Mạnh, et al. “Nghiên cứu thuật toán phân loại phương tiện giao thông dựa trên thị giác máy tính.”
- [5] Tuấn, Đ. H. A., & Thắng, N. M. (2021). Ứng dụng mô hình học sâu trong xác định các giai đoạn sinh trưởng của cây dưa lưới trồng trong nhà màng. *Bản B của Tạp chí Khoa học và Công nghệ Việt Nam*, 63(11).
- [6] H. Hoashi, T. Joutou, and K. Yanai. Image recognition of 85 food categories by feature fusion. In IEEE ISM, pages 296–301, 2010.
- [7] Kagaya, Hokuto, Kiyoharu Aizawa, and Makoto Ogawa. “Food detection and recognition using convolutional neural network.” In Proceedings of the 22nd ACM International Conference on Multimedia, pp. 1085–1088. 2014.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105
- [9] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” in European conference on computer vision. Springer, 2014, pp. 346–361

[10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems, 2015, pp. 91–99.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

10.3. Danh mục các công trình đã công bố thuộc lĩnh vực của đề tài của chủ nhiệm và những thành viên tham gia nghiên cứu

- a) Của chủ nhiệm đề tài: Không
- b) Của các thành viên tham gia nghiên cứu: Không

11. TÍNH CÁP THIẾT CỦA ĐỀ TÀI

Vùng đất tây Nam Bộ là nơi mang một sắc thái cổ điển, ở đó chứa đựng tinh hoa về ẩm thực và văn hóa của dân tộc Việt Nam, nó cũng là vùng đất của những lễ hội truyền thống dân tộc. Trong những lần đi du lịch hoặc lễ hội,... cụ thể là ở miền Tây Nam Bộ. Chúng tôi nhận thấy còn rất nhiều người họ không biết về các món đặc sản ở vùng miền này, họ vào một quán ăn nào đó mà muốn gọi món đó mà không biết nó tên gì gây ra trở ngại. Chính vì thế, chúng tôi xây dựng một ứng dụng di động hỗ trợ quản bá món ăn, đồng thời cũng hỗ trợ giúp nhận dạng món ăn đặc sản của từng tỉnh thành miền Tây nam Bộ, từ đó sẽ giúp cho những người khách du lịch, lễ hội,... có trải nghiệm tốt hơn về chuyến đi. Hiện tại, chúng tôi chưa tìm được các ứng dụng di động có tích hợp nhận diện món ăn Tây Nam Bộ nào. Do đó, việc xây dựng ứng dụng di động để giới thiệu các món ăn đặc trưng của Miền Tây Nam Bộ và cho phép người dùng nhận diện một số món ăn đặc trưng của Vùng là vô cùng cần thiết.

12. MỤC TIÊU ĐỀ TÀI

Mục tiêu đề tài là xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ, và cho phép nhận diện một số món ăn đặc trưng của miền Tây Nam Bộ.

13. ĐỐI TƯỢNG, PHẠM VI NGHIÊN CỨU

13.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài ứng dụng di động và mô hình Yolov5

13.2. Phạm vi nghiên cứu

Một số món ăn đặc trưng ở khu vực miền Tây Nam Bộ.

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

14. CÁCH TIẾP CẬN, PHƯƠNG PHÁP NGHIÊN CỨU

14.1. Cách tiếp cận

Dựa vào tài liệu tham khảo. Sau đó, tiến hành xây dựng, thử nghiệm mô hình YoloV5. Và cuối cùng, đưa vào ứng dụng mô hình.

14.2. Phương pháp nghiên cứu

- Nghiên cứu lý thuyết từ các tạp chí, các bài báo, website ở lĩnh vực có liên quan
- Đề xuất phương pháp giải quyết bài toán, thực hiện triển khai và đánh giá kết quả đạt được

15. NỘI DUNG NGHIÊN CỨU VÀ TIẾN ĐỘ THỰC HIỆN

15.1. Nội dung nghiên cứu

- Nghiên cứu lý thuyết:
 - + Tìm hiểu về nền tảng xây dựng ứng dụng di động
 - + Tìm hiểu huấn luyện mô hình trên YoloV5
- Thu thập dữ liệu:
 - + Thu thập hình ảnh về các món ăn Tây Nam Bộ và xây dựng các nhãn tương ứng với hình ảnh
- Xây dựng ứng dụng
 - + Huấn luyện dữ liệu bằng mô hình YoloV5
 - + Xây dựng tập theo TensorFlow Lite để đưa vào ứng dụng di động
 - + Tích hợp mô hình nhận diện đối tượng vào ứng dụng di động

15.2. Tiến độ thực hiện

STT	Các nội dung, công việc thực hiện	Sản phẩm	Thời gian (bắt đầu-kết thúc)	Người thực hiện và số ngày thực hiện
1.	Nghiên cứu lý thuyết về các mô hình để giúp nhận diện đối tượng trong ảnh (đặc biệt mô hình YoloV5) và các nghiên cứu liên quan đến ứng dụng di động	Khoảng 10-12 trang A4 trình bày báo cáo lý thuyết về các mô hình giúp nhận diện đối tượng, YoloV5 và phương pháp xây dựng ứng dụng di động dựa trên mô hình.	6/2022 – 7/2022	Nguyễn Duy Khang (5 ngày) Nguyễn Thị Mỹ Khánh (5 ngày) Nguyễn Hiếu Nghĩa (5 ngày) Lê Hải Yến (5 ngày) Tống Phước Lộc (5 ngày)

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

2.	Thu nhập và xây dựng tập dữ liệu mô hình	Khoảng 2-3 trang A4 trình bày về tập dữ liệu thu thập.	7/2022 – 9/2022	Nguyễn Duy Khang (9 ngày) Nguyễn Thị Mỹ Khanh (9 ngày) Nguyễn Hiếu Nghĩa (9 ngày) Lê Hải Yến (9 ngày) Tống Phước Lộc (9 ngày)
3.	Huấn luyện dữ liệu thu được dựa trên mô hình YoloV5	Khoảng 5-8 trang trình bày về phương pháp thực hiện để xây dựng.	9/2022 – 10/2022	Nguyễn Duy Khang (7 ngày) Nguyễn Thị Mỹ Khanh (7 ngày)
4.	Xây dựng ứng dụng di động theo TensorFlow Lite	Khoảng 5-8 trang trình bày về phương pháp thực hiện để xây dựng.	9/2022 – 10/2022	Nguyễn Hiếu Nghĩa (7 ngày) Lê Hải Yến (7 ngày) Tống Phước Lộc (7 ngày)
5.	Kiểm thử và đánh giá	Khoảng 3-5 trang A4 trình bày về phương pháp kiểm thử, đánh giá kết quả đạt được.	11/2022	Nguyễn Duy Khang (4 ngày) Nguyễn Thị Mỹ Khanh (4 ngày) Tống Phước Lộc (4 ngày)
6.	Viết báo cáo tổng kết	Quyển báo cáo tổng kết.	11/2022	Nguyễn Hiếu Nghĩa (4 ngày) Lê Hải Yến (4 ngày)

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

16. SẢN PHẨM			
Số thứ tự	Tên sản phẩm	Số lượng	Yêu cầu chất lượng sản phẩm (mô tả chi tiết chất lượng sản phẩm đạt được như nội dung, hình thức, các chỉ tiêu, thông số kỹ thuật,...)
I	Sản phẩm khoa học (Các công trình khoa học sẽ được công bố: sách, bài báo khoa học...): Không		
II	Sản phẩm đào tạo (Luận văn tốt nghiệp đại học): Không		
III	Sản phẩm ứng dụng: ứng dụng di động nhận diện ẩm thực miền Tây Nam Bộ		

17. PHƯƠNG THỨC CHUYÊN GIAO KẾT QUẢ NGHIÊN CỨU VÀ ĐỊA CHỈ ỨNG DỤNG			
17.1. Phương thức chuyển giao			
Tài liệu liên quan xây dựng ứng dụng di động và nhận diện món ăn			
17.2. Địa chỉ ứng dụng			
Khoa Công nghệ Thông tin và Truyền thông, trường Đại học Cần Thơ			
18. TÁC ĐỘNG VÀ LỢI ÍCH MANG LẠI CỦA KẾT QUẢ NGHIÊN CỨU			
18.1. Đối với lĩnh vực giáo dục và đào tạo			
Truyền tải văn hóa ẩm thực miền Tây Nam Bộ, hỗ trợ người nghiên cứu về ẩm thực và du lịch Việt Nam.			
18.2. Đối với lĩnh vực khoa học và công nghệ có liên quan			
Thúc đẩy sinh viên có thêm động lực làm nghiên cứu khoa học, góp phần làm đa dạng các công trình nghiên cứu của Trường Đại học Cần Thơ.			
18.3. Đối với phát triển kinh tế-xã hội			
Khi ứng dụng đưa vào thực tiễn sẽ mang lại một số lợi ích nhất định cho người dùng: tìm kiếm và cập nhật thông tin món ăn nhanh hơn. Thúc đẩy quảng bá văn hóa ẩm thực và đặc sản miền Tây Nam Bộ ra thế giới. Ứng dụng sẽ là cầu nối quan trọng thúc đẩy phát triển ngành du lịch nước nhà.			
18.4. Đối với tổ chức chủ trì và các cơ sở ứng dụng kết quả nghiên cứu			
Kết quả nghiên cứu của đề tài là hình thức minh chứng, quảng bá cho chất lượng giảng dạy, cũng là góp phần làm phong phú số lượng đề tài nghiên cứu cho Khoa CNTT&TT trường Đại học Cần Thơ. Hơn nữa, đề tài là cơ sở để Khoa hỗ trợ trực tiếp cho ngành du lịch địa phương cũng như nước nhà.			

Đề tài: Xây dựng ứng dụng di động giới thiệu ẩm thực miền Tây Nam Bộ

19. KINH PHÍ THỰC HIỆN ĐỀ TÀI VÀ NGUỒN KINH PHÍ

Kinh phí thực hiện đề tài: 15.000.000 đồng.

Trong đó:

Kinh phí Trường cấp: 15.000.000 đồng.

Các nguồn khác: 0 đồng.

Đơn vị tính: đồng

Số	Khoản chi, nội dung chi	Tổng kinh phí	Nguồn kinh phí	
			Kinh phí Trường cấp	Các nguồn khác
1	Chi mua vật tư, nguyên, nhiên, vật liệu	0	0	0
2	Chi tiền công lao động trực tiếp	12.275.000	12.275.000	0
3	Chi văn phòng, phẩm, thông tin liên lạc, in ấn	0	0	0
4	Chi hợp đồng đánh giá, nghiệm thu	2.725.000	2.725.000	0
Tổng cộng		15.000.000	15.000.000	0

Ngày 01 tháng 06 năm 2022

KHOA CNTT&TT

CÁN BỘ HƯỚNG DẪN

CHỦ NHIỆM ĐỀ TÀI

Nguyễn Hữu Hòa

Lâm Nhuet Khang

Nguyễn Phi Mỹ Khanh

TL.HIỆU TRƯỞNG
TRƯỞNG PHÒNG QUẢN LÝ KHOA HỌC



Lê Nguyễn Đoan Khôi