

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN CUỐI KỲ

ANJITA

ỨNG DỤNG GỢI Ý CÔNG THỨC NẤU ĂN TỪ HÌNH ẢNH NGUYÊN LIỆU

Môn học: Ứng dụng thị giác máy tính

Lớp: 19_23

Giảng viên: TS. Võ Hoài Việt

Thành phố Hồ Chí Minh, ngày 30 tháng 12 năm 2022

1. Mục lục

1. Mục lục	2
2. Thông tin thành viên nhóm.....	3
3. Các chức năng.....	3
4. Giao diện web và hướng dẫn sử dụng	3
4.1. Giao diện desktop.....	3
4.2. Giao diện mobile.....	6
5. Khảo sát các phương pháp.....	8
5.1. Yolov5.....	9
5.2. Google Vision API.....	9
5.3. Facebook inverse cooking.....	10
6. Inverse Cooking - Recipe generation	11
6.1. Gợi ý công thức nấu ăn từ hình ảnh.....	11
6.2. Tối ưu.....	12
7. Một số kết quả mô hình	12
8. Deploy.....	15
9. Future works	15
10. Tài liệu tham khảo	15

2. Thông tin thành viên nhóm

STT	MSSV	Tên	Ghi chú
1	19120450	Nguyễn Quốc Anh	
2	19120456	Nguyễn Phan Quốc Bảo	

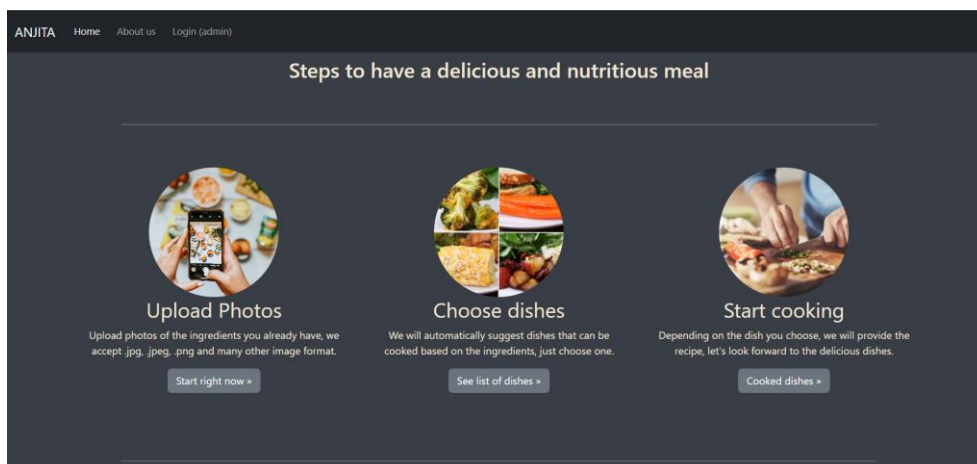
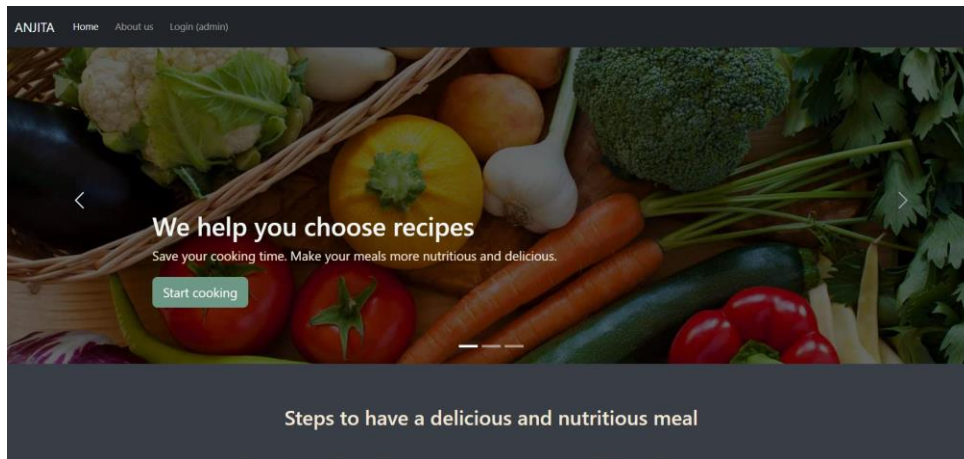
3. Các chức năng

STT	Chức năng	Hoàn thành	Ghi chú
1	Login (admin)	100%	Username: admin Pass: 0987654321
2	Upload hình ảnh	100%	
3	Nhận diện nguyên liệu nấu ăn (rau, củ, quả, thịt) có trong ảnh	100%	
4	Truy xuất database và đưa ra gợi ý món ăn, công thức nấu ăn	100%	
5	Thêm công thức nấu ăn (admin)	0	
Project https://anjita-bpnkcyfhmq-as.a.run.app		90%	

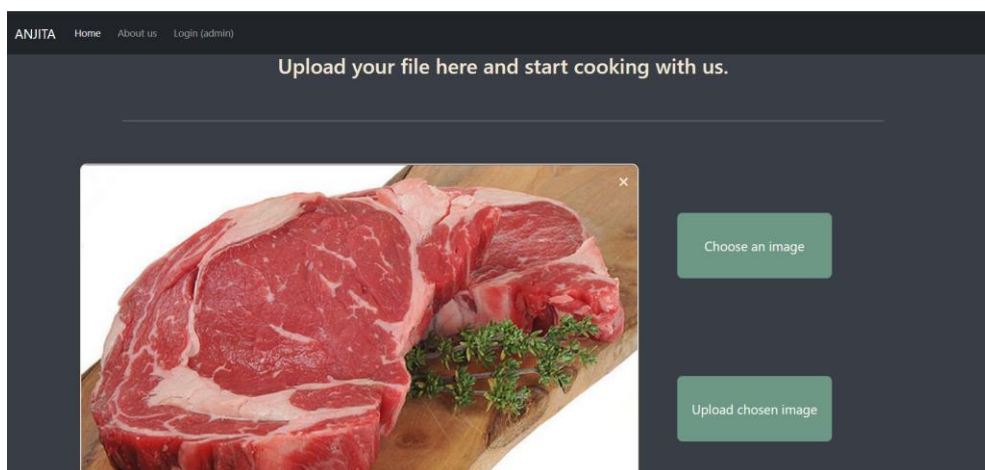
4. Giao diện web và hướng dẫn sử dụng

4.1. Giao diện desktop

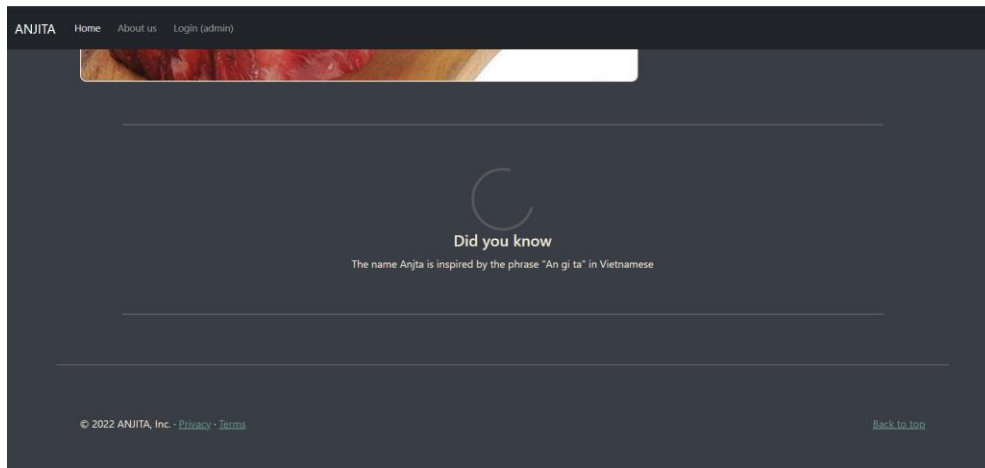
Khi truy cập link, Home page sẽ có dạng:



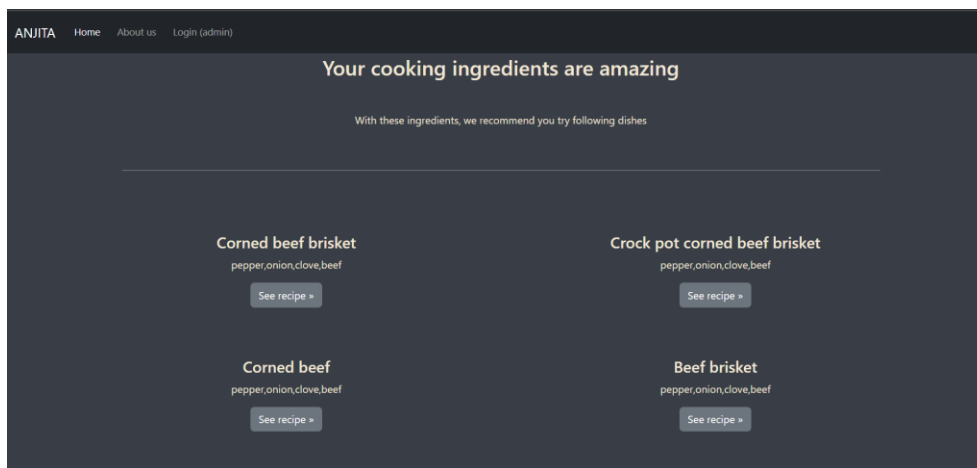
Lướt đến box upload hình ảnh hoặc nhấn vào start cooking để tự động cuộn. Hệ thống cho phép người dùng kiểm tra lại hình ảnh mình sử dụng để tránh chọn nhầm và tránh lộ thông tin. Ảnh chỉ được gửi đến cho hệ thống xử lý khi người dùng chọn Upload chosen image.



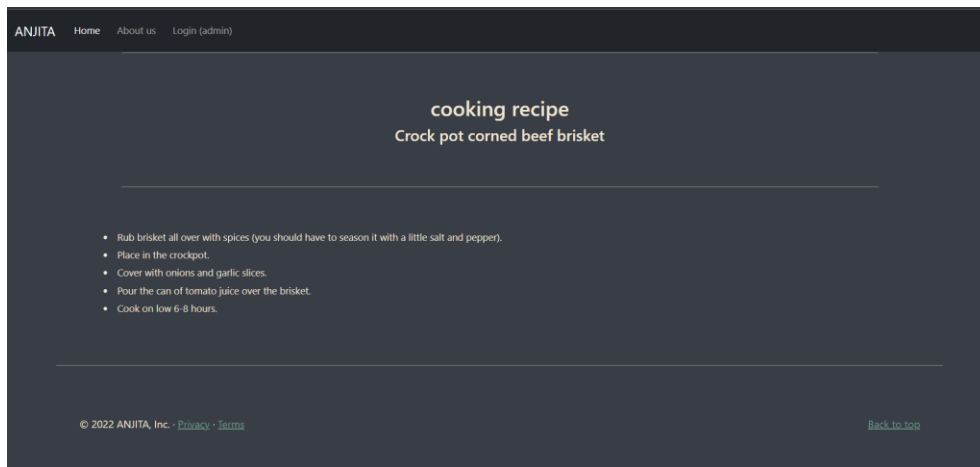
Giao diện loading chờ kết quả xử lý:



Giao diện hiển thị kết quả, danh sách các món ăn được tạo sinh, đồng thời cũng cung cấp tên những nguyên liệu cần thiết để tạo ra món ăn đó. Người dùng có thể chọn lựa món ăn phù hợp với mình để tham khảo công thức nấu ăn:



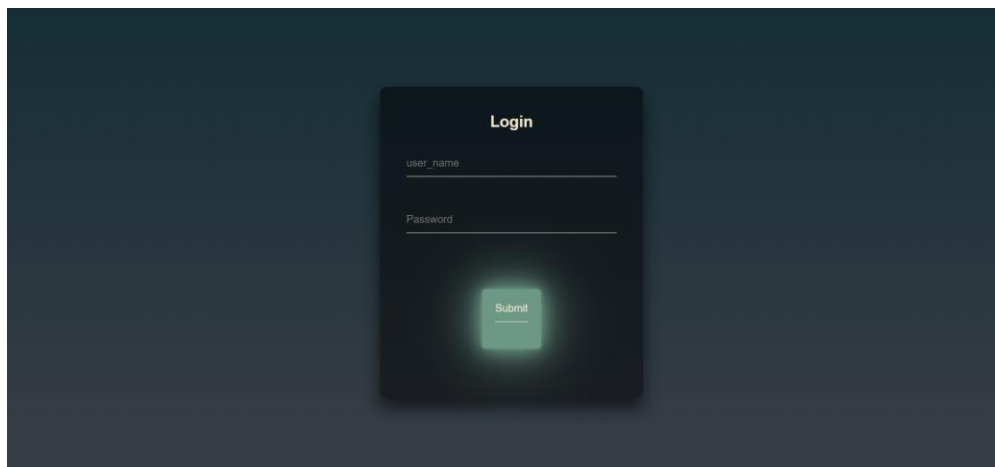
Giao diện hiển thị công thức nấu ăn:



About us page

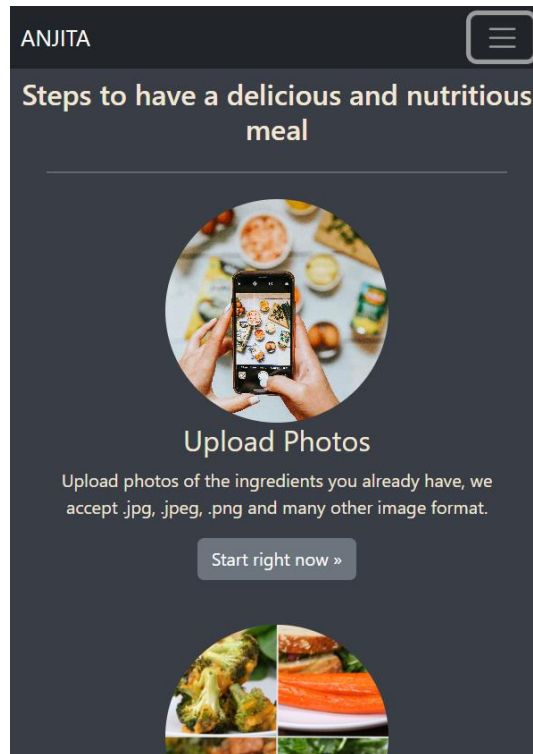
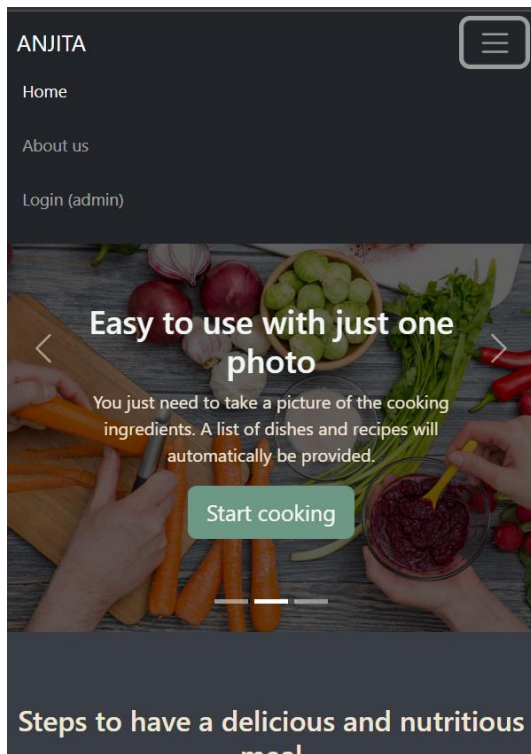


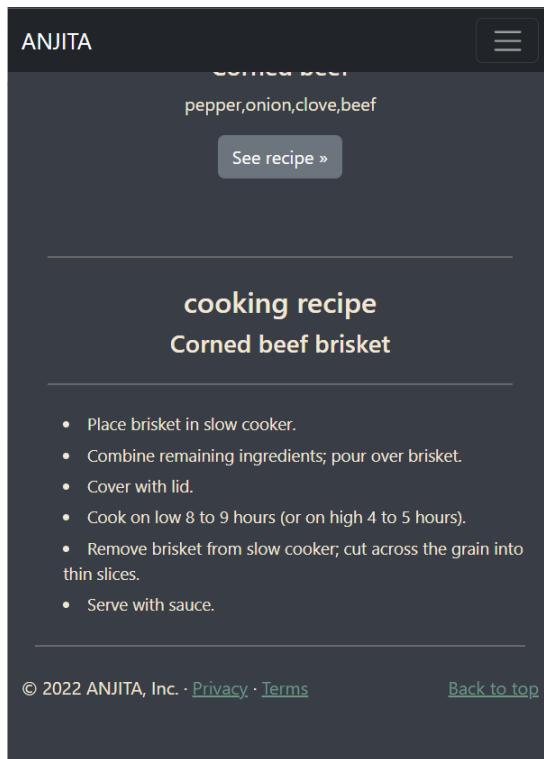
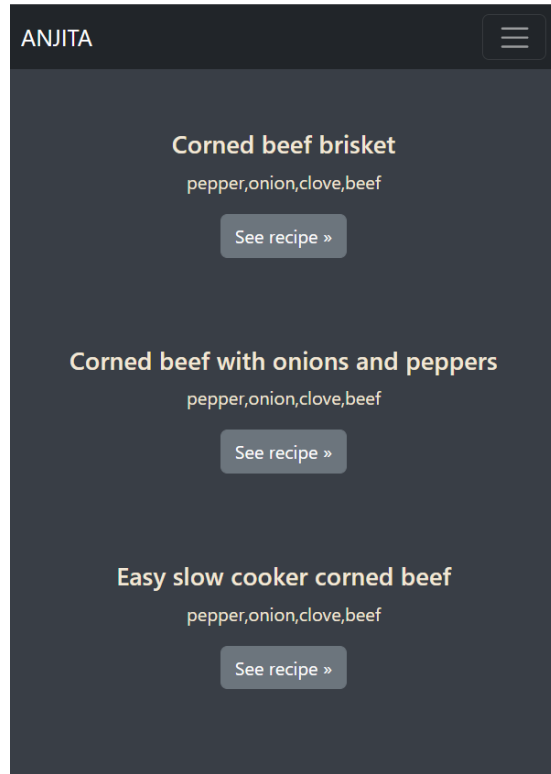
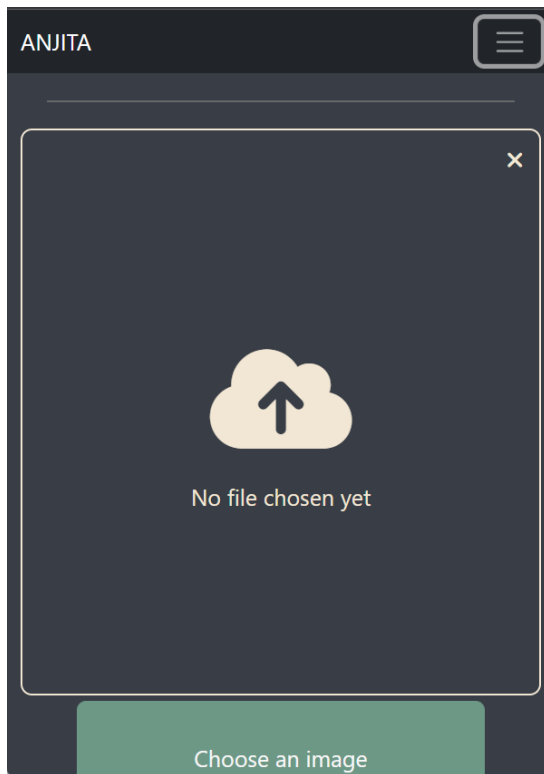
Login page:



4.2. Giao diện mobile

Web được thiết kế responsive để phù hợp với các kích thước màn hình khác nhau, giao diện cho các màn hình điện thoại:





5. Khảo sát các phương pháp

5.1.Yolov5

Sử dụng Yolov5 detect các nguyên liệu trong ảnh, sau đó tìm kiếm các công thức có xuất hiện tên các nguyên liệu

Vấn đề gặp phải:

- Các bộ trọng số có sẵn ko đủ mạnh mẽ để xác định nguyên liệu trong ảnh

Hướng giải quyết:

- Chia nhỏ ảnh thành các ô lưới và detect trên từng ảnh nhỏ, tìm kiếm các model segmentation để loại bỏ chi tiết thừa, các model phát hiện vùng quan tâm trong ảnh để tập trung vào những đối tượng này

Kết quả:

- Các open source segmentation không đủ chuyên hóa cho vấn đề tách thức ăn, làm cho kết quả tiền xử lý không tốt, việc chia nhỏ hình ảnh thành các ảnh con không tăng tính chính xác của model như dự đoán.

5.2.Google Vision API

Sử dụng vision api của google để tự động gán nhãn cho hình ảnh, từ đó thu thập tên các nguyên liệu

Vấn đề gặp phải:

- Vision api không được train chuyên biệt cho vấn đề nhận diện nguyên liệu
- Vision api có đến 20000 class trong đó chỉ có 257 class là tên các nguyên liệu nấu ăn (1%)
- Model thường trả về nhãn chung chung (ví dụ như Food, Ingredient, Sea Food,...)

Hướng cải thiện:

- Chia nhỏ ảnh thành lưới 3x3, 4x4 rồi gọi API cho từng ảnh kết hợp với ảnh lớn

Kết quả:

- Có cải thiện độ chính xác, tuy nhiên thời gian xử lý rất lâu và hoàn toàn không phù hợp khi triển khai do số lần gọi API tăng gấp 26 lần, tức chi phí phải trả cho google cũng tăng lên 26 lần.

5.3.Facebook inverse cooking

Sử dụng model inverse cooking của facebook, đây là một model hoàn toàn tự động end to end cho việc tạo sinh công thức từ hình ảnh. Model nhận input là ảnh, dự đoán các nguyên liệu và tự động sinh ra tên món ăn cũng như cách nấu từ danh sách các từ vựng mà hoàn toàn không truy vấn bất kì tập dữ liệu nào.

Kết quả:

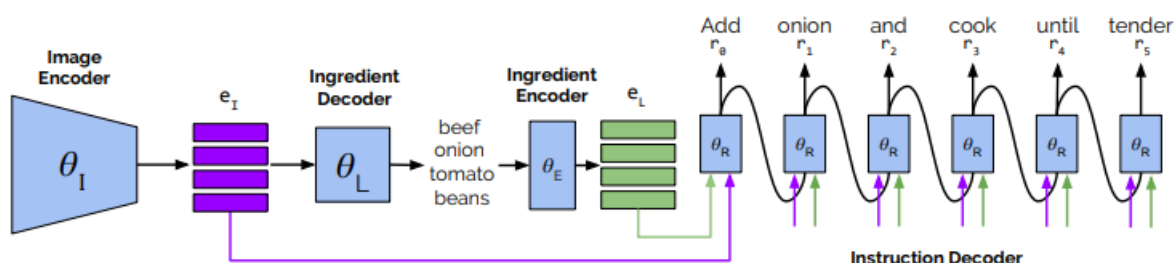
- Model hoạt động tốt do được huấn luyện chuyên biệt cho mục đích mà ứng dụng hướng tới.
- Open source, không tốn chi phí.

6. Inverse Cooking - Recipe generation

6.1. Gợi ý công thức nấu ăn từ hình ảnh

Tạo một công thức (tiêu đề, thành phần và hướng dẫn) từ một hình ảnh là một nhiệm vụ đầy thách thức, đòi hỏi sự hiểu biết đồng thời về các thành phần tạo nên món ăn cũng như quá trình biến đổi mà chúng đã trải qua, ví dụ: cắt lát, pha trộn hoặc trộn với các thành phần khác.

Quy trình tạo công thức có đầu vào là các nguyên liệu mà người dùng cung cấp thông qua ảnh. Sau đó, các nguyên liệu của món ăn sẽ được suy luận và chuỗi hướng dẫn sẽ được tạo tùy thuộc vào hình ảnh đầu vào và danh sách nguyên liệu tương ứng của nó để tạo ra món ăn.



Ảnh mô hình tạo công thức. Từ hình ảnh θ_I , tham số hóa và trích xuất các đặc trưng hình ảnh e_I bằng bộ mã hóa hình ảnh. Thành phần trong ảnh sẽ được dự đoán bởi θ_L và được mã hóa thành e_L bằng bộ mã hóa thành phần θ_E . Bộ giải mã hướng dẫn nấu ăn, được tham số hóa bởi θ_R , bộ giải mã này tạo ra một tiêu đề công thức và một chuỗi các bước nấu ăn bằng cách tham khảo các đặc trưng hình ảnh e_I , các đặc trưng thành phần e_L và các từ được dự đoán trước đó (r_0, \dots, r_{t-1}).

Hình ảnh trên mô tả cách tiếp cận bài toán. Hệ thống tạo công thức lấy hình ảnh thực phẩm làm đầu vào và xuất ra một chuỗi các hướng dẫn nấu ăn, được tạo bằng bộ giải mã lấy hai phần đặc trưng làm đầu vào. Phần đầu tiên đại diện cho các đặc trưng trực quan được trích xuất từ một hình ảnh, trong khi phần thứ hai mã hóa các nguyên liệu nấu ăn được trích xuất từ hình ảnh.

Nhóm tác giả sử dụng những kiến trúc mạng state of the art như resnet để rút trích đặc trưng ảnh và transformer cho các khối tạo sinh tên và công thức nấu ăn.

6.2. Tối ưu

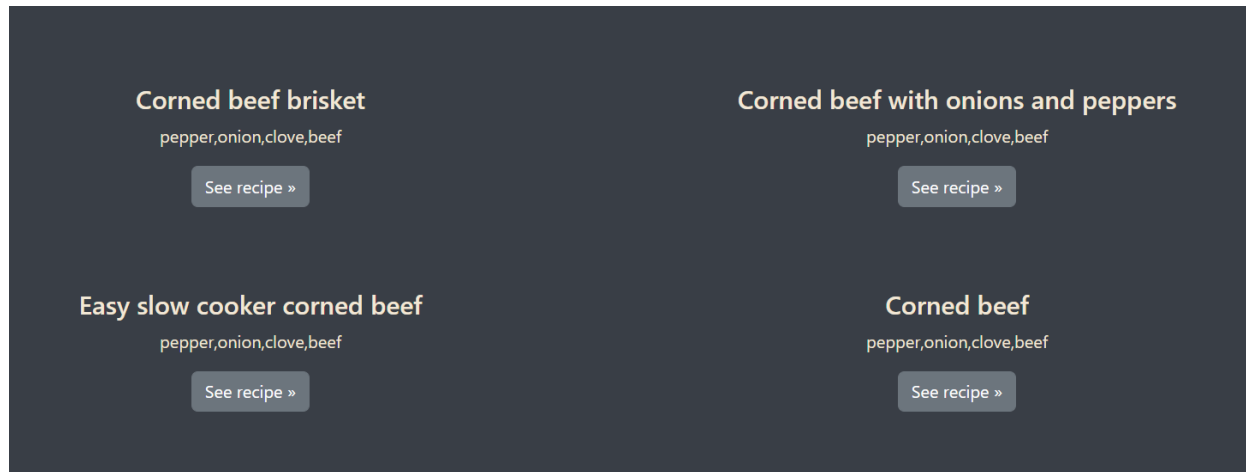
Nhóm tác giả huấn luyện mô hình recipe transformer theo 2 giai đoạn. Trong giai đoạn đầu tiên, mạng rút trích đặt trung hình ảnh và dự đoán các thành phần được huấn luyện trước. Sau đó, trong giai đoạn thứ hai, nhóm tác giả huấn luyện một mô hình mã hóa những nguyên liệu nấu ăn và một mô hình dự đoán các bước của công thức nấu ăn bằng cách tối ưu các hàm lỗi nhằm điều chỉnh các bộ tham số θ_R và θ_E . Trong khi huấn luyện, bộ giải mã các bước nấu ăn sẽ chỉ lấy đầu vào là các thành phần nguyên liệu.

7. Một số kết quả mô hình

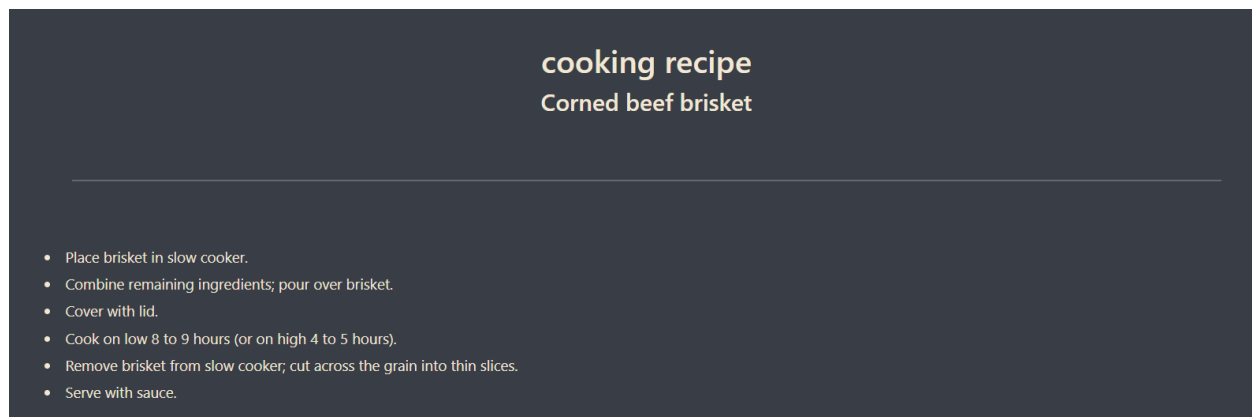
Input



Kết quả món ăn:



Kết quả công thức



Input:



Kết quả món ăn

Chimichurri sauce pepper,oil,salt,clave,onion,parsley See recipe »	Simple sauteed red pepper and onion pepper,oil,salt,clave,onion,parsley See recipe »
Parsley and garlic oil pepper,oil,salt,clave,onion,parsley See recipe »	Sauteed peppers pepper,oil,salt,clave,onion,parsley See recipe »

Kết quả công thức:

cooking recipe

Simple sauteed red pepper and onion

- Heat a large skillet over medium heat.
- Add the olive oil into the skillet and heat until shimmering.
- Add the onions and cook, stirring occasionally with a wooden spoon, until golden, about 10 to 15 minutes.
- Add the garlic and cook until the garlic is fragrant, about 30 seconds.
- Add the red pepper flakes and the parsley, and cook, stirring occasionally, for about 10 minutes.
- Season with sea salt.
- Serve immediately.

8. Deploy

Hệ thống được dockerize và deploy trên google cloud Run, các chi tiết cài đặt cho local có thể tham khảo file README.md đính kèm trong source code.

Link deploy: <https://anjita-bpnkcyfhmq-as.a.run.app/>

9. Future works

- Thêm các chức năng bổ dung từ vựng vào bộ từ điển có sẵn.
- Có thể train lại model để phù hợp hơn với nguyên liệu sống cũng như các món ăn Việt Nam.
- Xây dựng các test case kiểm tra hệ thống.

10. Tài liệu tham khảo

Inverse cooking: <https://arxiv.org/pdf/1812.06164.pdf>