NHẬN DẠNG CÁC LOẠI THỨC ĂN ĐƯỜNG PHỐ VIỆT NAM VÀ TÍNH CHẤT CỦA TỪNG MÓN ĂN

Sinh viên thực hiện: Nguyễn Văn Tâm
Trường Đại Học Sư Phạm Kỹ Thuật Thành Phố Hồ Chí Minh, Việt Nam, 2022

Tóm tắt - Thực ăn và tính chất thành phần dinh dưỡng đóng vai trò chính trong các vấn đề liên quan đến sức khỏe, nó ngày càng trở nên thiết yếu hơn trong cuộc sống hàng ngày của chúng ta. Trong bài báo này, áp dụng mạng nơ-ron CNN cho nhiệm vụ phát hiện và nhận dạng hình ảnh thực ăn đường phố ở Việt Nam. Với sự đa dạng của nhiều loại thực phẩm, việc nhận dạng hình ảnh của các loại thực phẩm thường vô cùng phức tạp. Các nhà nguyên cứu đã chứng minh là một kỹ thuật nhận dạng hình ảnh hữu ích và CNN là cách tốt để nhận dạng thông qua hình ảnh. Em đã áp dụng CNN cho việc phát hiện và nhận dạng thông qua tạo dựng model. Những điểm nổi bật mà Mạng nơ-ron Convolutions (CNN) học được đã được nhận thấy được chính xác và nhanh hơn những cách làm thủ công.

1.1.1. Giới thiệu

Thức ăn đường phố là các loại thức ăn, đồ uống được chế biến sẵn, các món này được phục vụ tại chỗ và là thức ăn nhanh, chi phí rẻ hơn một bữa ăn trong các nhà hàng và nhanh chóng, tiện lợi, giá cả phải chăng. Theo Tổ chức Lương thực và Nông nghiệp Liên Hợp Quốc (FAO) thì khoảng 2,5 tỷ người ăn thức ăn đường phố mỗi ngày. Thức ăn đường phố có mối liên hệ mật thiết Take-out, đồ ăn vặt (hàng rong, quà vặt), đồ ăn nhẹ (snack), thức ăn nhanh, nó được phân biệt bởi hương vị địa phương và được mua trên đường phố, mà không cần nhập bất kỳ trụ sở hay công trình xây dựng gì.

Trên thế giới đã có một số công trình nghiên về cứu nhận dạng các loại thức ăn bằng cách vận dụng kỹ thuật xử lý ảnh và nhận dạng đã đạt được một số kết quả khả quan. Các hệ thống nhận dạng được nhiều loại thức ăn và đồ uống, từ đó dự đón và gợi ý về chế độ dinh dưỡng cho người dùng, bảo đảm cung cấp đủ chất dinh dưỡng cho người dùng.

Nhận dạng thức ăn thông qua xử lý ảnh bằng thuật toán CNN, để nhận diện đúng loại và tính chất của món ăn, từ đó có thể đưa ra dự đoán chính xác về thành phần dinh dưỡng cung cấp cho người dùng. Vì vậy, em đã chọn nghiên cứu về đề tài "Nhận dạng thức ăn đường phố Việt Nam" một ứng dụng nhận dạng dựa vào hình ảnh mang tính chất của món ăn đó.

Thực phẩm và chế độ ăn uống để tránh xa các bệnh nhiễm trùng sắp xảy ra hoặc hiện có. Việc sắp xếp một ứng dụng để sàng lọc chế độ ăn uống của mọi người một cách tự nhiên, sẽ giúp ích trong nhiều khía cạnh sức khoẻ. Nó mở rộng sự chú ý đến các cá nhân về xu hướng thực phẩm và chế độ ăn uống của họ. Trong hai mươi năm gần đây nhất, nghiên cứu đã tập trung vào việc nhận thức thực phẩm và dữ liệu lành mạnh của

chúng từ các bức ảnh được chụp bằng cách sử dụng các quy trình AI và thị giác. Để khảo sát thích hợp việc chấp nhận chế độ ăn uống, việc đánh giá chính xác ước tính lượng calo của thực phẩm có ý nghĩa chính. Một phần lớn các cá nhân đang say mê và không đủ năng động. Với mức độ bận rộn và tập trung vào các cá nhân ngày nay, thật dễ dàng bỏ qua việc theo dõi thực phẩm họ ăn. Tương tự như vậy, xem xét tình hình nhốt và cô lập hiện tại trong đại dịch COVID-19 trên diện rộng, nhiều cá nhân có xu hướng ngấu nghiến và không đối phó với chế độ ăn uống của họ. Điều này duy nhất xây dựng tầm quan trọng của việc phân loại thực phẩm thích hợp.

Bài toán của chúng ta bao gồm việc phân loại hình ảnh thực phẩm thành ba lớp khác nhau. Có hai thách thức chính trong nhiệm vụ này. Vấn đề cơ bản là cùng một loại thực phẩm được chế biến khác nhau tùy thuộc vào địa điểm, nguyên liệu sẵn có và sở thích cá nhân của người nấu.

Một cái khác có thể là góc mà hình ảnh được chụp. Những thách thức này gây ra sự khác biệt đáng kể trong các hình ảnh từ cùng một lớp và làm cho vấn đề phân loại khó khăn hơn.

Chúng tôi chọn phân loại hình ảnh thực phẩm vì mục đích của chúng tôi là tạo ra một chương trình có thể được sử dụng trong các ứng dụng đánh giá chế độ ăn uống. Trong bài báo này, chúng tôi đã áp dụng Mạng nơ-ron tích hợp (CNN) để giải quyết vấn đề này. Chúng tôi sẽ giới thiệu việc sử dụng mô hình của chúng tôi giống như chu trình chuẩn bị và kết quả.

Trong bài báo này, một nỗ lực đã được thực hiện để mô tả các hình ảnh của các mặt hàng thực phẩm cho các ứng dụng quan sát thói quen ăn uống bổ sung sử dụng mạng nơ-ron phức hợp (CNN). Vì CNN được trang bị để xử lý nhiều thông tin và có thể đánh giá các tính năng một cách tự nhiên, chúng đã được sử dụng để thực hiện việc phân loại thực phẩm. Bộ dữ liệu Food-101 tiêu chuẩn đã được chọn làm cơ sở thông tin hoạt động cho dự án này.

2. Phương pháp

2.1. Deep learning

Deep learning được bắt nguồn từ thuật toán **Neural network** vốn xuất phát chỉ là một ngành nhỏ của Machine Learning. Deep Learning là một chi của ngành máy học dựa trên một tập hợp các thuật toán để cố gắng mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý với cấu trúc phức tạp, hoặc bằng cách khác bao gồm nhiều biến đổi phi tuyến.

Tương tự như cách chúng ta học hỏi từ kinh nghiệm thuật toán, deep learning sẽ thực hiện một nhiệm vụ nhiều lần mỗi lần tinh chỉnh nhiệm vụ một chút để cải thiện

kết quả. Deep Learning chỉ đơn giản là kết nối dữ liệu giữa tất cả các tế bào thần kinh nhân tạo và điều chỉnh chúng theo dữ liệu mẫu.

2.2. Mạng nơ-ron CNN

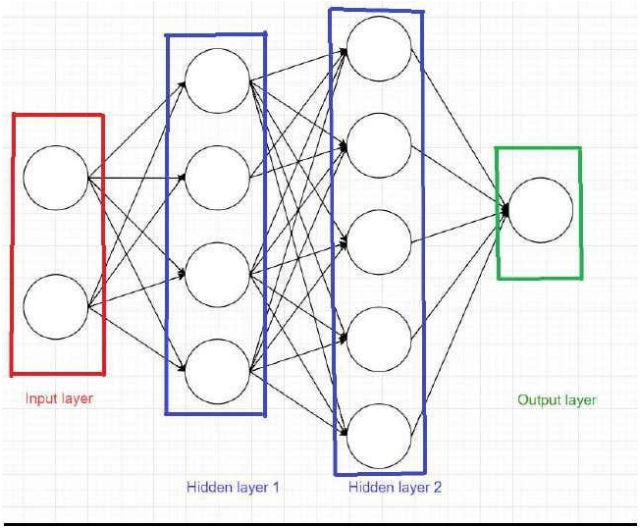
CNN đưa ra một phương pháp tiên tiến để nhận dạng hình ảnh, Nó là một tổ chức thần kinh đa lớp. Mạng nơron tích chập (CNN) là một thuật toán Deep Learning có thể lấy hình ảnh đầu vào, gán độ quan trọng các đối tượng khác nhau trong hình ảnh và có thể phân biệt được từng đối tượng này với nhau. Công việc tiền xử lý được yêu cầu cho mạng nơron tích chập thì ít hơn nhiều so với các thuật toán phân loại khác. Trong các phương thức sơ khai, các bộ lọc được thiết kế bằng tay (hand - engineered), với một quá trình huấn luyện để chọn ra các đặc trưng phù hợp thì mạng nơron tích chập lại có khả năng tự học để chọn ra các đặc trưng tối ưu nhất.

Kiến trúc của nơron tích chập tương tự như mô hình kết nối của các nơron trong bộ não con người và được lấy cảm hứng từ hệ thống vỏ thị giác trong bộ não (visual cortex). Các nơ-ron riêng lẻ chỉ phản ứng với các kích thích trong một khu vực hạn chế của trường thị giác được gọi là Trường tiếp nhận (Receptive Field). Một tập hợp các trường như vậy chồng lên nhau để bao phủ toàn bộ khu vực thị giác.

2.3. Cấu trúc của mạng CNN

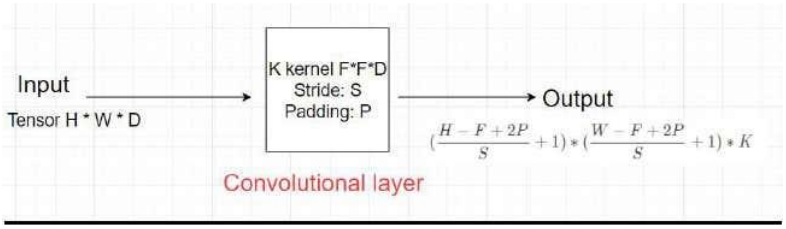
2.3.1. Convolutional layer

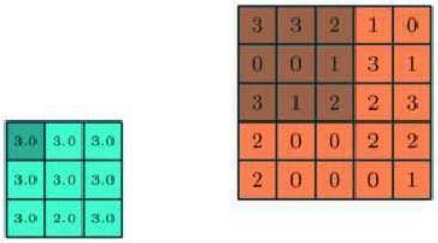
Mỗi hidden layer được gọi là fully connected layer, tên gọi theo đúng ý nghĩa, mỗi node trong hidden layer được kết nối với tất cả các node trong layer trước. Cả mô hình được gọi là fully connected neural network (FCN).

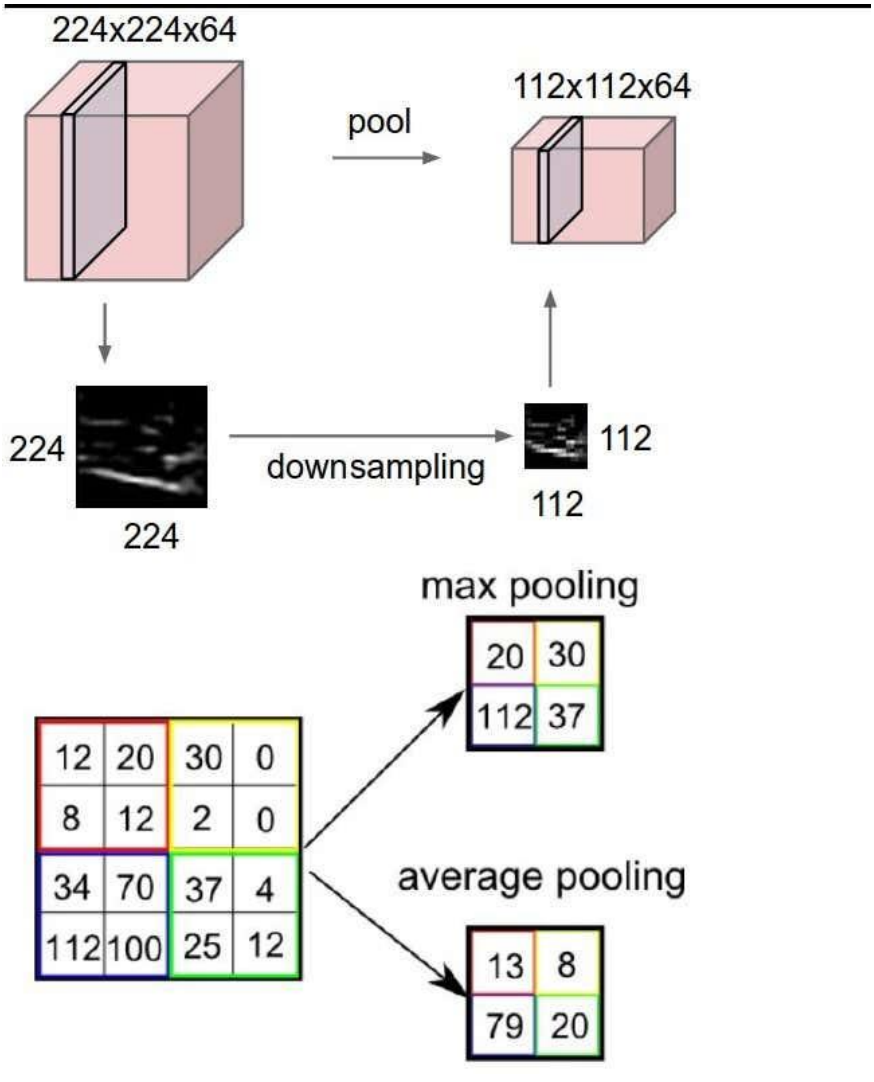


Hình 1: Convolutional layer

Pooling layer và Fully connected layer

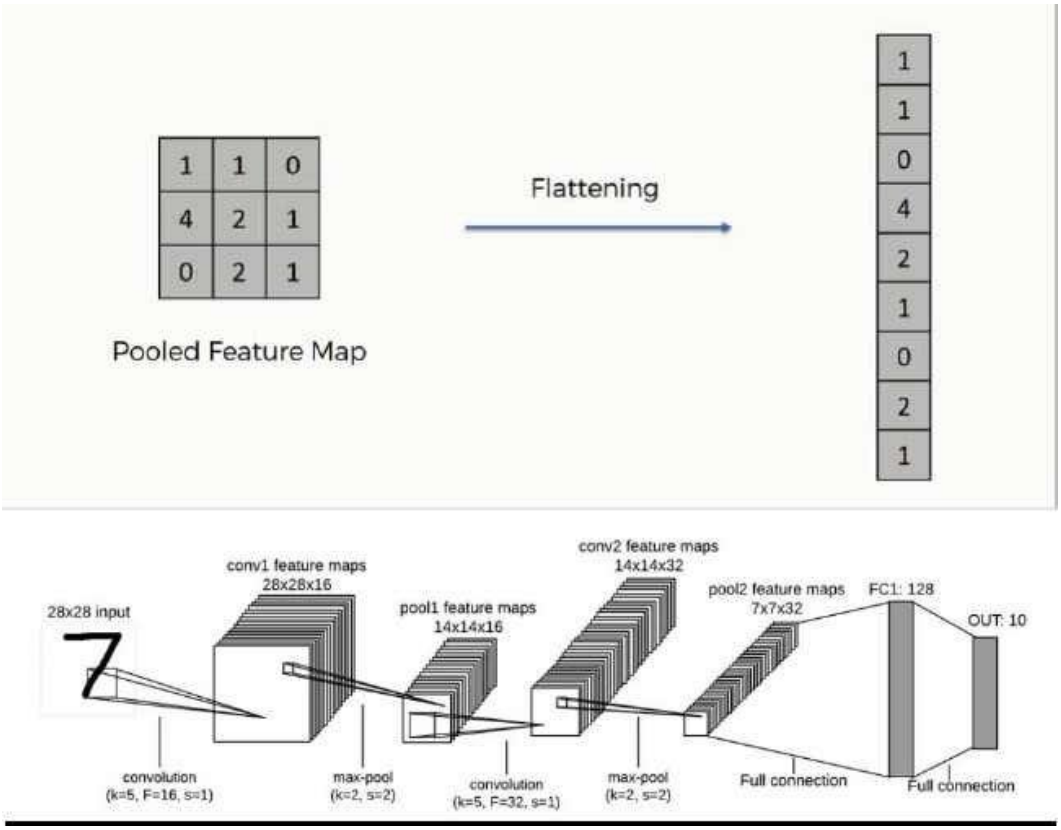






Hình 2: Pooling layer và Fully connected layer

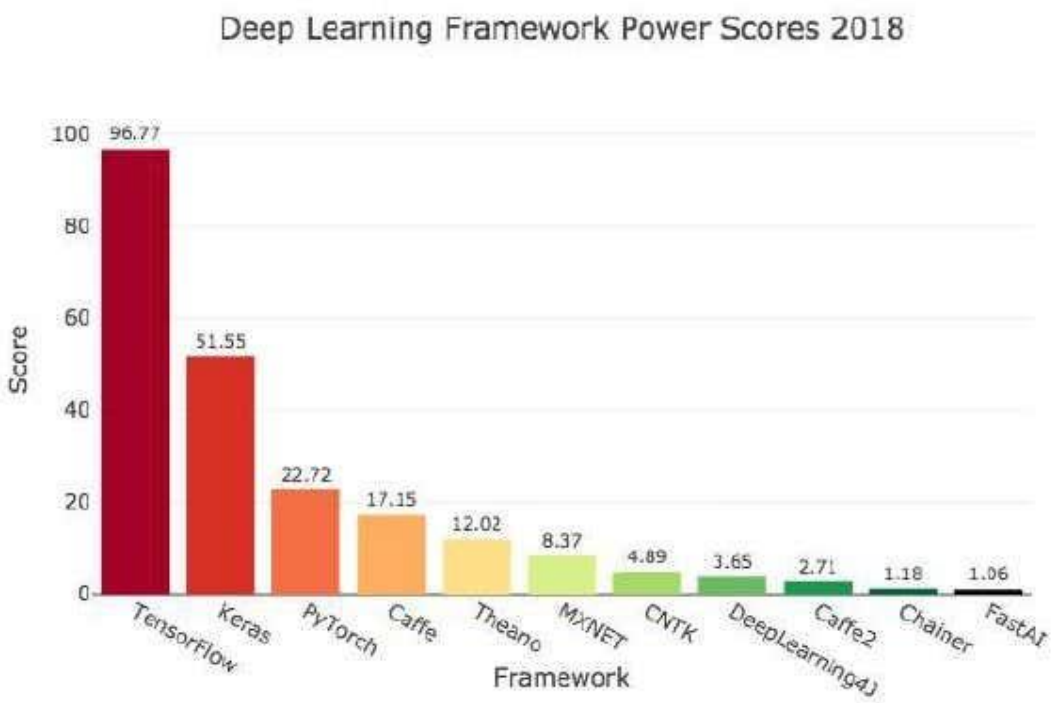
Sau khi ảnh được truyền qua nhiều convolutional layer và pooling layer thì model đã học được tương đối các đặc điểm của ảnh (ví dụ mắt, mũi, khung mặt, ...) thì tensor của output của layer cuối cùng, kích thước H*W*D, sẽ được chuyển về 1 vector kích thước (H*W*D)



Hình 3: Fully connected layer

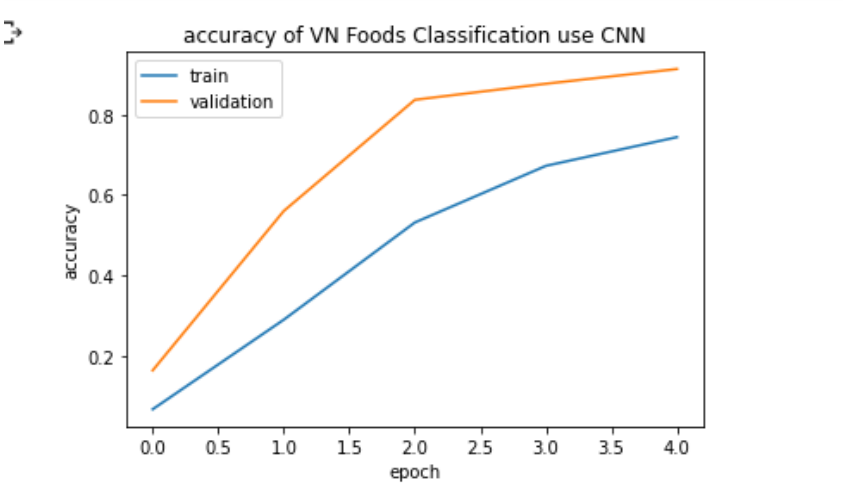
2.3.2. Keras

Framework mã nguồn mở cho deep learning được viết bằng Python. Nó có thể chạy trên nền của các deep learning framework khác như: tensorflow, theano, CNTK. Với các API bậc cao, dễ sử dụng, dễ mở rộng, keras giúp người dùng xây dựng các deep learning model một cách đơn giản



Hình 4: Deep learnng framework

3. Kết Quả

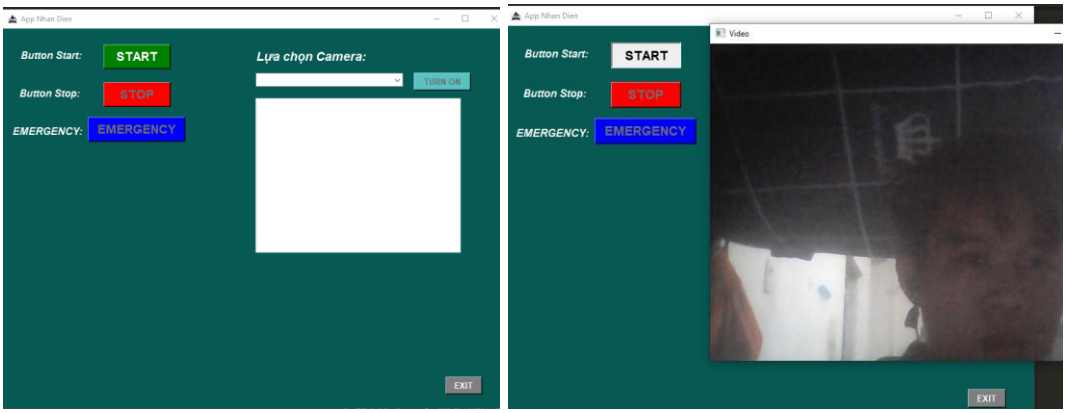


Hình 5: Độ chính xác của quá trình training

Hình 6: Nhận diện thức ăn bằng hình ảnh

Hình 7: Nhận diện realtime bằng colab

Hình 8: Giao diện load dữ liệu



Hình 9: Giao diện app

4. Kết Luận

Đề tài này cho thấy rằng mạng nơ-ron tích chập có thể giải quyết thành công các vấn đề phân loại hình ảnh thức ăn với số lượng lớp tương đối nhỏ. Việc phân loại thành nhiều lớp hơn đòi hỏi nhiều kiến trúc phức tạp hơn. Bên cạnh độ phức tạp của một mô hình, việc chọn các chức năng tối ưu hóa và các tham số phù hợp cũng rất quan trọng. Với bộ dữ liệu đã cho cho 15 loại thức ăn: độ chính xác cuối cùng của mô hình đạt 90%.

Tài liệu tham khảo

[1] Lowe, D.G. Object Recognition from Local Scale Invariant Features. In Proceedings of the ICCV'99, Corfu, Greece, 20–21 September 1999.

[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[3] David Poole, "Artificial intelligence – Foundations of computational agents". 2010 [Online]. Available: http://artint.info/html/ArtInt_180.html [Accessed: Sept. 28, 2015].

[4] BLVC, "The toolkit for the CNN – Caffe" 2014. [Online]. Available: http://caffe.berkeleyvision.org/ [Accessed: Sept. 28, 2015]

[5] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).