



ĐH SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH
KHOA CƠ KHÍ CHẾ TẠO MÁY

Trần Sỹ Việt 20146544

Phát hiện quả cà chua sử dụng mạng Convolutional Neural Networks(CNN)

<i>Thông tin</i>	<i>Tóm tắt</i>
<i>Trường đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh, Khoa Cơ khí Chế tạo máy. Tháng 5/2023</i>	Bài báo này trình bày về việc sử dụng mạng Neural Convolutional (CNN) để phát hiện quả cà chua trong hình ảnh. Quá trình huấn luyện và đánh giá mô hình CNN được thực hiện trên nền tảng Colab, một môi trường trực tuyến mạnh mẽ và thuận tiện để xây dựng và chạy các mô hình học sâu. Chúng tôi xây dựng một mô hình CNN có khả năng phát hiện quả cà chua với độ chính xác cao và đã tiến hành huấn luyện và đánh giá mô hình trên tập dữ liệu được gán nhãn đúng. Trên cơ sở đó, có thể phát triển mô hình vào trong các hệ thống phân loại hoa quả dựa trên các đặc tính của nó bằng việc áp dụng trí tuệ nhân tạo vào ứng dụng mô hình sản xuất tiên tiến.
<i>Từ khóa</i> <i>Tomato Detection CNN AI</i>	

1. Giới thiệu.

Phát hiện và phân loại đối tượng trong hình ảnh là một lĩnh vực quan trọng trong thị giác máy tính. Trong bài báo này, chúng tôi tập trung vào việc phát hiện quả cà chua trong hình ảnh. Việc phát hiện quả cà chua có thể áp dụng rộng rãi trong các lĩnh vực như nông nghiệp tự động, công nghệ thực phẩm và thống kê về sản lượng cây trồng. Mạng Neural Convolutional (CNN) là một kiến trúc mạng thị giác phổ biến được sử dụng để xử lý hình ảnh và đạt được nhiều kết quả ấn tượng trong việc phân loại và phát hiện đối tượng.

2. Phương pháp

CNN là một kiến trúc mạng thị giác được xây dựng trên cơ sở của các lớp tích chập và lớp kết nối đầy đủ. Các lớp tích chập trong CNN được sử dụng để tìm kiếm các đặc trưng cục bộ trong hình ảnh, trong khi các lớp kết nối đầy đủ sẽ tổng hợp các đặc trưng này và đưa ra dự đoán cuối cùng. Kiến trúc CNN đã chứng tỏ hiệu suất cao trong việc phân loại và phát hiện đối tượng trong hình ảnh. Do vậy để mục đích của nghiên cứu để phát hiện và nhận biết quả cà chua bằng mạng nơ-ron tích chập CNN.

3. Mô hình và thuật toán

3.1 Mô hình.

Bộ dữ liệu của nghiên cứu được thu thập trực tiếp hình ảnh trên google bao gồm các hình ảnh chụp quả cà chua và các loại quả không phải là cà chua. Hình ảnh bao gồm các trạng thái khác nhau của quả chua, gồm màu sắc, giống cây, kích thước,.. Bộ dữ liệu dùng để huấn luyện bao gồm ba thư mục: train, test, validation, trong đó có thư mục quả cà chua với số lượng gần 500 hình và hơn 1000 hình các loài quả khác cà chua ví dụ như: chuối, táo, lê, ...

3.2 Thuật toán

- Mạng nơ ron tích chập Convolutional Neural Networks (CNN)

Mạng neural tích chập (CNN) là một trong những kiến trúc mạng neural phổ biến nhất trong lĩnh vực xử lý ảnh và nhận dạng đối tượng. Với kiến trúc được thiết kế đặc biệt để xử lý các hình ảnh và đặc trưng liên quan đến không gian, CNN đã đạt được những thành công đáng kể trong việc giải quyết các bài toán phân loại, nhận dạng và phát hiện đối tượng.

CNN được xây dựng dựa trên các lớp tích chập (convolutional layer) để trích xuất đặc trưng của hình ảnh. Sau đó, các lớp liên kết đầy đủ (fully connected layer) được sử dụng để phân loại ảnh. Kiến trúc này cho phép mô hình học các đặc trưng quan trọng của hình ảnh, giúp cho việc phân loại, nhận dạng và phát hiện đối tượng trở nên chính xác hơn.

Trong thời gian gần đây, CNN đã được ứng dụng rộng rãi trong các lĩnh vực khác như xử lý âm thanh, nhận dạng giọng nói, và xử lý ngôn ngữ tự nhiên. Một số ứng dụng của CNN bao gồm nhận dạng khuôn mặt, phân loại sản phẩm và xe cộ, giúp cải thiện hiệu quả và độ chính xác của các hệ thống tự động.

- **Tăng cường dữ liệu**

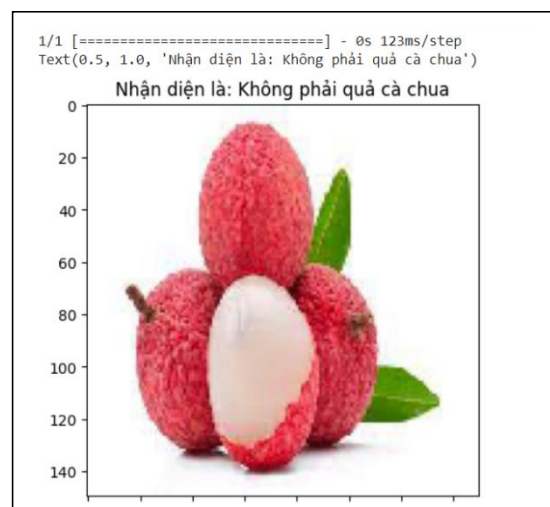
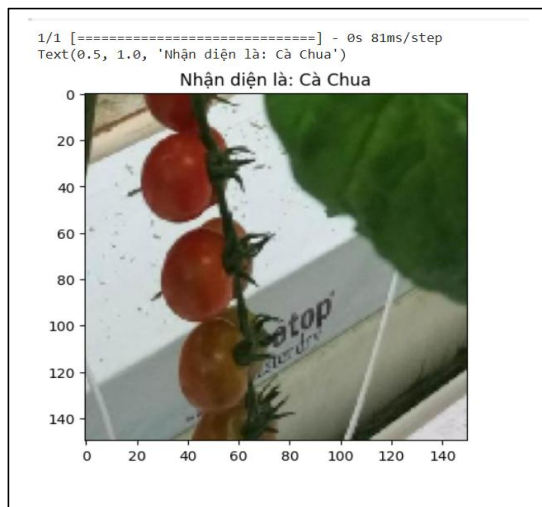
Tăng cường dữ liệu (data augmentation) là một kỹ thuật quan trọng trong việc huấn luyện mạng neural tích chập (CNN) để cải thiện hiệu suất và khả năng tổng quát hóa của mô hình. Khi dữ liệu huấn luyện hạn chế, việc tạo ra các phiên bản biến đổi của dữ liệu gốc có thể giúp mô hình hiểu và nhận diện các đặc trưng quan trọng hơn.

Có nhiều kỹ thuật tăng cường dữ liệu phổ biến được áp dụng trong CNN. Một trong số đó là phép xoay (rotation), trong đó hình ảnh được quay một góc nhất định để tạo ra các góc nhìn khác nhau. Thứ hai, phép phóng to/thu nhỏ (scaling) cũng là một kỹ thuật tăng cường dữ liệu phổ biến. Kỹ thuật tiếp theo là phép dịch chuyển (translation), trong đó hình ảnh được di chuyển theo các hướng khác nhau. Ngoài ra, các kỹ thuật tăng cường dữ liệu khác bao gồm phép lật ảnh (flipping), thêm nhiễu (noise), cắt tỉa (cropping) và biến đổi màu sắc (color augmentation). Tuy kỹ thuật tăng cường dữ liệu có thể tăng thời gian huấn luyện, nhưng nó đem lại lợi ích đáng kể bằng cách cải thiện độ chính xác và khả năng tổng quát hóa của mô hình CNN. Với sự tăng cường dữ liệu thông minh và phù hợp, chúng ta có thể đạt được mô hình CNN với hiệu suất cao và khả năng.

4. Kết quả

Mô hình CNN đã đưa ra chuẩn đoán chính xác việc nhận diện quả cà chua và có độ chính xác cao.

Kết quả sau khi thử nghiệm dự đoán trên colab:



Kết quả chạy trên web app xây dựng bằng thư viện Flask.

