HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

PHẠM NGỌC HOÀN

NGHIÊN CỨU MÔ HÌNH HỌC SÂU VÀ ỨNG DỤNG BIGDL CHO BÀI TOÁN NHẬN DIỆN VÀ PHÂN LOẠI NÔNG SẢN

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH

Mã số: 8.48.01.01

TÓM TẮT LUẬN VĂN THẠC SỸ

(Theo định hướng ứng dụng)

Hà Nội – 2023

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: PGS.TS. NGUYỄN VĂN THUỶ

Phản biện 1:

Phản biện 2:

Luận văn này được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc:

Có thể tìm hiểu luận văn này tại:

Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Học sâu có ứng dụng sâu rộng trong các lĩnh vực của đời sống như tìm kiếm sự khác nhau giữa các văn bản, phát hiện gian lận, phát hiện spam, nhận dạng chữ viết, giọng nói, nhận dạng hình ảnh,… góp phần quan trọng trong việc hỗ trợ con người trong nhiều lĩnh vực đời sống. Từ những ứng dụng thực tế và những lợi ích mà Học sâu đem lại, đề tài nghiên cứu "*Nghiên cứu mô hình học sâu và ứng dụng BIGDL cho bài toán nhận diện và phân loại nông sản* " đã được đưa ra với hy vọng có thể ứng dụng thành công các mô hình học sâu hiện đại để xây dựng một hệ thống nhận diện nông sản tự động.

2. Tổng quan về vấn đề nghiên cứu

Nhận diện vật thể trong ảnh được coi là bài toán cơ bản nhất trong lĩnh vực Thị giác máy tính, là nền tảng cho rất nhiều bài toán mở rộng khác như bài toán phân lớp, định vị, tách biệt vật thể. Tuy bài toán cơ bản này đã tồn tại hàng thế kỷ nhưng con người vẫn chưa thể giải quyết nó một cách triệt để, do tồn tại rất nhiều khó khăn để máy tính có thể hiểu được các thông tin trong một bức ảnh: sự đa dạng trong điểm nhìn, đa dạng trong kích thước, các điều kiện khác nhau của ánh sáng, sự lộn xộn phức tạp của nền,…

Với những ưu điểm trên, Đề tài nghiên cứu lựa chọn mô hình CNNs áp dụng cho bài toán nhận diện và phân loại nông sản thông qua mã nguồn mở BIGDL.

3. Mục đích nghiên cứu

Đề tài tìm hiểu ứng dụng nhận diện và phân loại nông sản cũng như cách triển khai công cụ tìm kiếm hình ảnh phần mềm tự động để giảm nguồn nhân lực và đảm bảo chất lượng phần hơn với công việc tìm kiếm bằng tay.

Mục tiêu chính của đề tài là nghiên cứu mô hình học sâu và ứng dụng nhận diện và phân loại nông sản để đạt được tốc độ tìm kiếm nhanh và chuẩn xác nhất để cho người dùng không mất nhiều thời gian tìm kiếm sản phẩm.

- Nghiên cứu về các hệ thống nhận diện hình ảnh.
- Thử nghiệm, đánh giá độ hiệu quả của các thuật toán.

- Xây dựng hệ thống nhận diện và phân loại nông sản tự động.

4. Đối tượng và phạm vi nghiên cứu

➢ Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài là mô hình học sâu và ứng dụng được mã nguồn mở BIGDL cho bài toán nhận diện và phân loại nông sản.

➢ Phạm vi nghiên cứu

- Số lượng nông sản sẽ nhận diện: 40 loại nông sản phổ biến ở nước ta như nho, táo, chuối, thanh long…

- Số lượng ảnh gốc cho mỗi loại quả: 500 ảnh, bao gồm các ảnh chụp nông sản ở các góc độ khác nhau với nền tùy ý, có thể lấy từ nguồn trên mạng hoặc tự chụp bằng thiết bị camera cá nhân.

5. Phương pháp nghiên cứu

- Phương pháp nghiên cứu lý thuyết

+ Đọc và phân tích tài liệu về các phương pháp, thuật toán đã từng được sử dụng để xây dựng hệ thống nhận diện hình ảnh.

- Phương pháp thực nghiệm

+ Thử nghiệm và đánh giá độ hiệu quả của các thuật toán.

+ Xây dựng hệ thống nhận diện hình ảnh

CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN

1.1 Bài toán nhận diện và phân loại nông sản

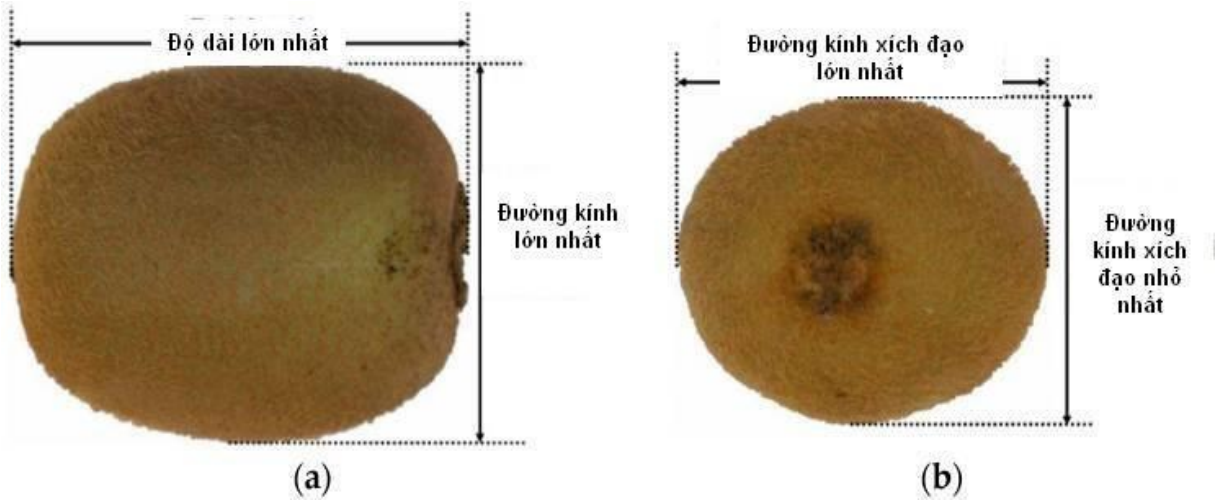
Nhận dạng vật thể trong ảnh được coi là bài toán cơ bản nhất trong lĩnh vực Thị giác máy tính, là nền tảng cho rất nhiều bài toán mở rộng khác như bài toán phân lớp, định vị, tách biệt vật thể.... Tuy bài toán cơ bản này đã tồn tại hàng thế kỷ nhưng con người vẫn chưa thể giải quyết nó một cách triệt để, do tồn tại rất nhiều khó khăn để máy tính có thể hiểu được các thông tin trong một bức ảnh.

Là một trường hợp cụ thể của bài toán nhận dạng và phân lớp, bài toán nhận dạng nông sản kế thừa các khó khăn vốn có của bài toán gốc, và kèm theo là các khó khăn riêng của chính nó, như: đa dạng về đối tượng; quy mô và độ phức tạp; đặc trưng và thuộc tính; ứng dụng trong ngành nông nghiệp.

Tổng quan, bài toán nhận diện và phân loại nông sản đặt ra những thách thức riêng và yêu cầu kiến thức về nông nghiệp, công nghệ thông tin và ứng dụng trong lĩnh vực nông nghiệp. Dữ liệu đầu vào và đầu ra của bài toán nhận diện và phân loại nông sản có thể khá đa dạng, tùy thuộc vào phạm vi và mục đích cụ thể của bài toán.

1.2 Các hướng tiếp cận và giải quyết bài toán

Bài toán tự động nhận dạng nông sản đã xuất hiện từ lâu và đã có rất nhiều bài báo, công trình khoa học được đưa ra nhằm đề xuất hoặc cải tiến các thuật toán nhận dạng. Trong đó, xuất hiện sớm nhất là các phương pháp Xử lý ảnh – Image Processing, các phương pháp này tập trung vào phát triển các thuật toán nhằm trích xuất thông tin, ví dụ các tham số về màu sắc, hình dạng, kết cấu, kích thước…, từ bức ảnh đầu vào để nhận dạng nông sản [2, 3]. Do chỉ đơn thuần xử lý trên một vài ảnh đầu vào trong khi sự biến thiên về màu sắc, hình dạng, kích thước… của nông sản quá phức tạp, kết quả đạt được của các phương pháp này không được cao và phạm vi áp dụng trên số lượng loại nông sản cũng bị hạn chế.



Hình 1.1: Các thông tin về hình học được tính toán bởi các thuật toán Xử lý ảnh

1.2.1 Phương pháp Học máy truyền thống

Học máy là một lĩnh vực của trí tuệ nhân tạo, nghiên cứu về việc phát triển các thuật toán và mô hình để giúp máy tính có khả năng học hỏi và cải thiện hiệu suất trong việc giải quyết các vấn đề. Học máy có thể được áp dụng để giải quyết các bài toán trong nhiều lĩnh vực khác nhau như thị giác máy tính, xử lý ngôn ngữ tự nhiên, truy vấn thông tin, điều khiển robot và phân tích dữ liệu.

1.2.1.1. Trích chọn đặc trưng

Trích chọn đặc trưng (Feature Engineering hoặc Feature Extraction) là quá trình lựa chọn và trích xuất các đặc trưng quan trọng và phù hợp nhất để mô tả dữ liệu đầu vào trong quá trình học máy. Các đặc trưng này có thể là các thông tin trực tiếp từ dữ liệu như độ dài, chiều rộng, màu sắc, cường độ, hoặc được tạo ra thông qua các kỹ thuật phân tích dữ liệu phức tạp hơn như PCA (Principal Component Analysis) và LDA (Linear Discriminant Analysis).

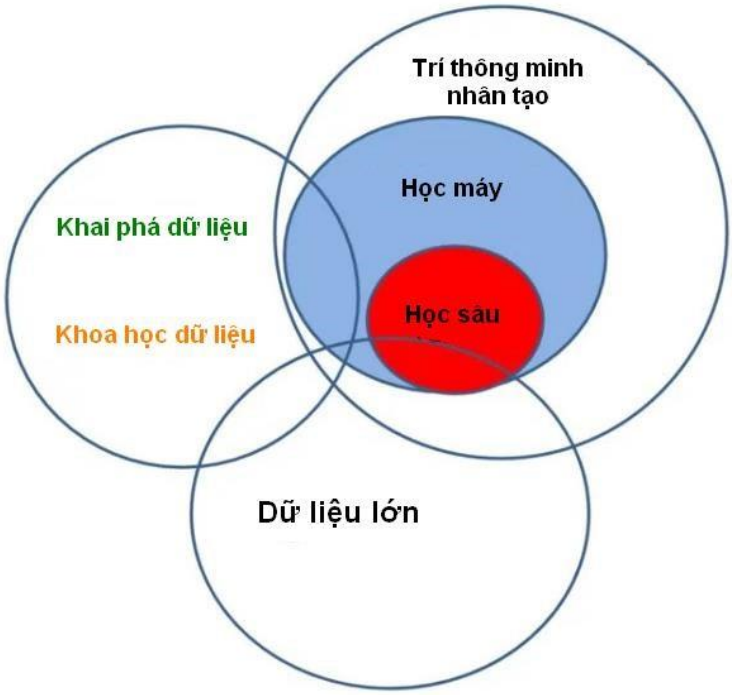
1.2.1.2. Thuật toán

Thuật toán phân loại là một trong những thuật toán cơ bản của học máy, nó được sử dụng để phân loại dữ liệu vào các nhóm khác nhau dựa trên các đặc trưng của chúng.

1.2.2 Phương pháp Học sâu

Học sâu (deep learning) là một lĩnh vực của trí tuệ nhân tạo (AI) liên quan đến việc sử dụng một mạng lưới nơ-ron nhân tạo (artificial neural network) để học và trích xuất các đặc trưng từ dữ liệu đầu vào. Phương pháp học sâu đã đạt được nhiều thành

công đáng kể trong các lĩnh vực như xử lý ngôn ngữ tự nhiên, thị giác máy tính và nhận dạng giọng nói.



Hình 1.2: Mối quan hệ của Học sâu với các lĩnh vực liên quan

Các phương pháp học sâu bao gồm nhiều lớp nơ-ron được kết nối với nhau để tạo thành một mạng lưới nơ-ron sâu(Deep Neural Network). Mỗi lớp nơ-ron đóng vai trò trích xuất các đặc trưng từ dữ liệu đầu vào, và các lớp này được kết hợp với nhau để tạo ra một mô hình học sâu có khả năng tự động học và cải thiện (xem Hình 1.6).

Các phương pháp học sâu thường được huấn luyện thông qua một quá trình tối ưu hóa tham số, ví dụ như sử dụng thuật toán lan truyền ngược (backpropagation) để điều chỉnh trọng số của các liên kết giữa các nơ-ron trong mạng lưới.

Hình 1.3: Bức ảnh quả tạ hai đầu sinh ra bởi mô hình dự đoán Học sâu

1.3 Thành tựu của phương pháp Học sâu trong các lĩnh vực

Học sâu (deep learning) đã đạt được nhiều thành tựu quan trọng trong các lĩnh vực khác nhau, bao gồm:

Phân tích ngữ nghĩa văn bản

Y học

Tài chính

Tự động lái xe

Thị giác máy tính

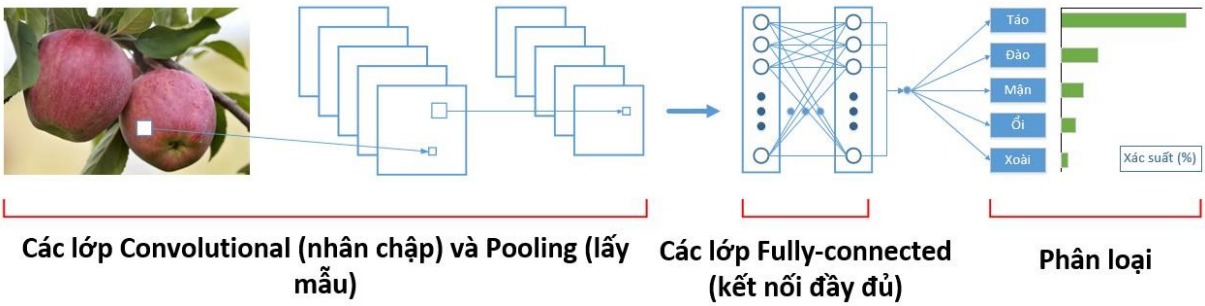
1.4 Kết luận chương

Như đã trình bày trong phần mở đầu, mục đích của luận văn là tìm hiểu và ứng dụng một mô hình Học sâu vào bài toán nhận dạng, phân loại nông sản, nguyên nhân chính khiến Học sâu được chọn làm giải pháp là bởi khả năng mạnh mẽ vượt trội của nó đối với các phương pháp Học máy truyền thống khi áp dụng vào các bài toán nhận dạng vật thể, trong đó vật thể là các đối tượng rất khó chọn lọc đặc trưng phù hợp, cụ thể với trường hợp này là các nông sản. Để chứng minh cho nhận định này, luận văn đã thực hiện phép so sánh độ chính xác của hai mô hình nhận dạng, được huấn luyện lần lượt bởi hai phương pháp trên với cùng bộ dữ liệu đầu vào. Kết quả cụ thể sẽ được trình bày trong Chương 3 – Kết quả thực nghiệm và Đánh giá.

CHƯƠNG 2. PHƯƠNG PHÁP NHẬN DIỆN, PHÂN LOẠI NÔNG SẢN

2.1 Mô hình mạng nơron tích chập

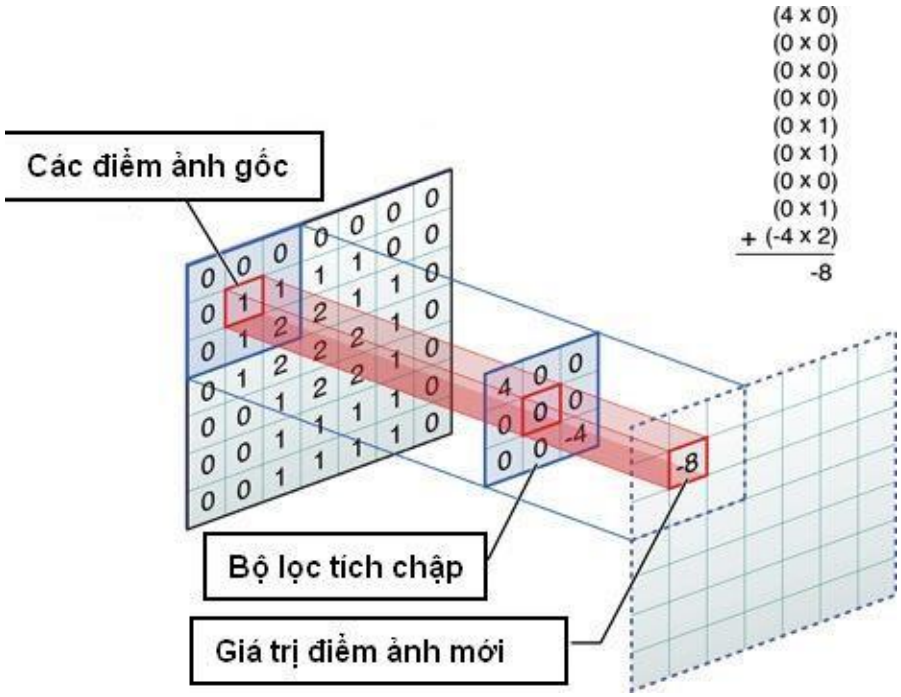
Các lớp cơ bản trong một mạng CNN bao gồm: Lớp tích chập (Convolutional), Lớp kích hoạt phi tuyến ReLU (Rectified Linear Unit), Lớp lấy mẫu (Pooling) và Lớp kết nối đầy đủ (Fully-connected). Trong một số trường hợp, các lớp này có thể được xếp chồng lên nhau để tạo thành một kiến trúc mạng phức tạp hơn. Ví dụ, một mô hình CNN thông thường có thể bao gồm nhiều lớp tích chập, lớp kích hoạt và lớp tổng hợp, trước khi kết thúc bằng các lớp kết nối đầy đủ và đầu ra.



Hình 2.1: Kiến trúc cơ bản của một mạng tích chập

- Lớp tích chập:

Lớp tích chập (convolutional layer) là một lớp quan trọng trong kiến trúc của mạng nơ-ron tích chập (CNN). Lớp tích chập giúp mạng CNN trích xuất các đặc trưng từ dữ liệu đầu vào bằng cách sử dụng các bộ lọc (filters) để quét (convolve) qua các vùng của dữ liệu.



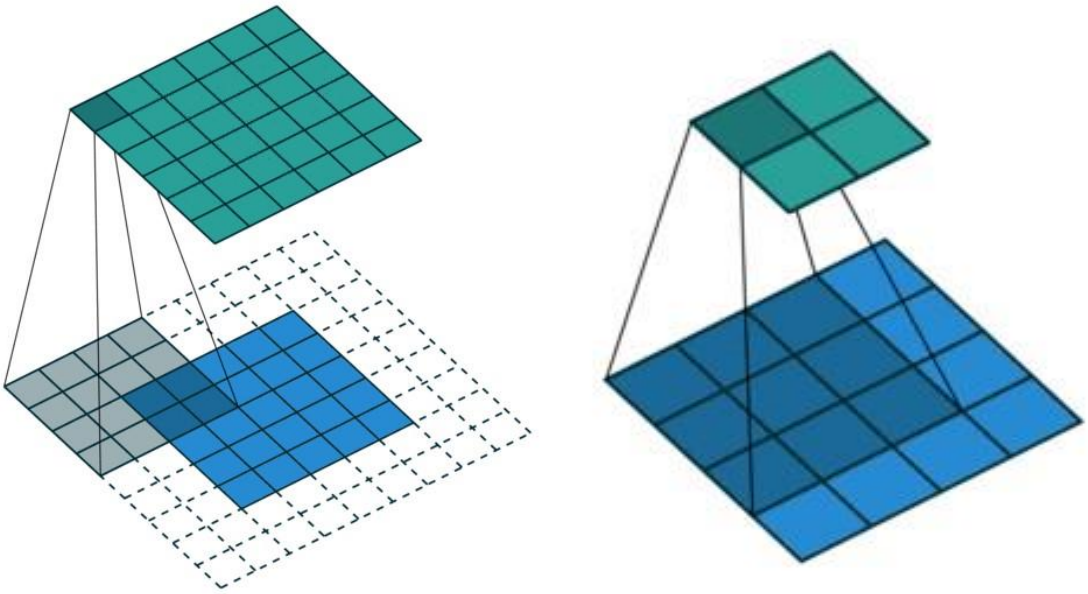
Hình 2.2: Ví dụ bộ lọc tích chập được sử dụng trên ma trận điểm ảnh

Trong ví dụ ở Hình 2.2 [12], bộ lọc được sử dụng là một ma trận có kích thước 3x3. Bộ lọc này được dịch chuyển lần lượt qua từng vùng ảnh đến khi hoàn thành quét toàn bộ bức ảnh, tạo ra một bức ảnh mới có kích thước nhỏ hơn hoặc bằng với kích thước ảnh đầu vào. Kích thước này được quyết định tùy theo kích thước các khoảng trắng được thêm ở viền bức ảnh gốc và được tính theo công thức (1) [13]:

$$o = \frac{i+2*p-k}{s} + 1 \quad (1)$$

Trong đó:

- o: kích thước ảnh đầu ra
- i: kích thước ảnh đầu vào
- p: kích thước khoảng trắng phía ngoài viền của ảnh gốc
- k: kích thước bộ lọc
- s: bước trượt của bộ lọc



Hình 2.3: Trường hợp thêm/không thêm viền trắng vào ảnh khi tích chập

Như vậy, sau khi đưa một bức ảnh đầu vào cho lớp Tích chập ta nhận được kết quả đầu ra là một loạt ảnh tương ứng với các bộ lọc đã được sử dụng để thực hiện phép tích chập. Các trọng số của các bộ lọc này được khởi tạo ngẫu nhiên trong lần đầu tiên và sẽ được cải thiện dần xuyên suốt quá trình huấn luyện.

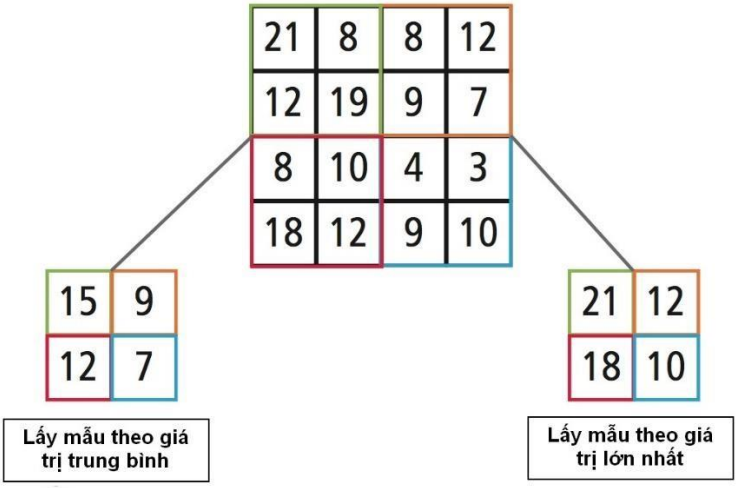
- **Lớp kích hoạt phi tuyến ReLU**:

Lớp kích hoạt phi tuyến ReLU (Rectified Linear Unit) là một lớp quan trọng trong các mạng nơ-ron, đặc biệt là trong mạng nơ-ron tích chập (CNN). ReLU được sử dụng để kích hoạt các nơ-ron trong các lớp tích chập và kết nối đầy đủ của mạng. Hàm kích hoạt ReLU được định nghĩa bằng cách đặt đầu ra của các nơ-ron có giá trị nhỏ hơn 0 bằng 0, và giữ nguyên giá trị đầu ra của các nơ-ron có giá trị lớn hơn hoặc bằng 0. Cụ thể, hàm ReLU được định nghĩa như sau:

$$f(x) = max(0, x) \quad (2)$$

- **Lớp lấy mẫu**:

Lớp lấy mẫu (pooling layer) là một lớp quan trọng trong mạng nơ-ron tích chập (CNN). Lớp này thường được sử dụng để giảm kích thước của bản đồ đặc trưng (feature map) và giảm thiểu overfitting.



Hình 2.4: Phương thức Avarage Pooling và Max Pooling

Như vậy, với mỗi ảnh đầu vào được đưa qua lấy mẫu ta thu được một ảnh đầu ra tương ứng, có kích thước giảm xuống đáng kể nhưng vẫn giữ được các đặc trưng cần thiết cho quá trình tính toán sau này.

- Lớp kết nối đầy đủ:

Lớp kết nối đầy đủ này được thiết kế hoàn toàn tương tự như trong mạng nơ-ron truyền thống, tức là tất cả các điểm ảnh được kết nối đầy đủ với node trong lớp tiếp theo. Lớp này thường được sử dụng để kết nối các bản đồ đặc trưng (feature maps) từ lớp tích chập và lớp lấy mẫu với nhau để tạo thành một đầu ra.

2.2 Các mô hình CNN phổ biến

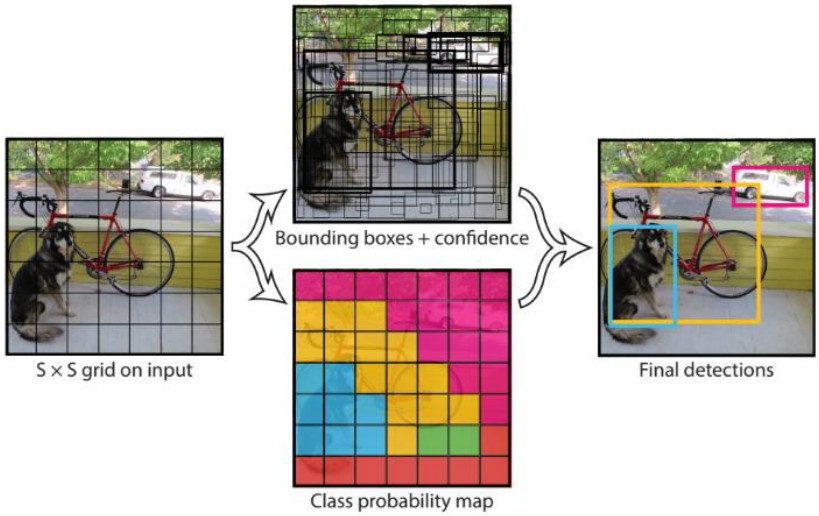
2.2.1 Faster R-CNN (2016)

Faster R-CNN (Faster Region-based Convolutional Neural Network) là một mô hình object detection (phát hiện đối tượng) phổ biến trong deep learning. Nó được giới thiệu bởi Shaoqing Ren, Kaiming He, Ross Girshick và Jian Sun vào năm 2015.

2.2.2 Lớp các mô hình họ YOLO

Yolo (You Only Look Once) là một họ các mô hình nhận diện đối tượng dựa trên deep learning, được phát triển bởi nhóm nghiên cứu của Joseph Redmon. Các mô hình Yolo được sử dụng rộng rãi trong các ứng dụng nhận diện đối tượng, như nhận diện giao thông, nhận diện khuôn mặt, và nhận diện vật thể trong ảnh chụp từ camera an ninh, camera giám sát.

YOLOv1 (2015)



Hình 2.5: Các bước xử lý trong mô hình YOLO [19]

YOLOv2 (2016)

2.2.3 SSD Model

SSD (Single Shot MultiBox Detector) là một thuật toán phát hiện đối tượng được sử dụng phổ biến trong lĩnh vực thị giác máy tính. Thuật toán này được giới thiệu bởi nhóm nghiên cứu của Google DeepMind vào năm 2016.

SSD sử dụng một mạng neural sâu để phân tích ảnh đầu vào và tạo ra các bounding box và xác suất lớp tương ứng cho mỗi đối tượng được phát hiện trong ảnh. Điều này được thực hiện bằng cách sử dụng các lớp tích chập bổ sung được thêm vào trên mạng cơ sở.

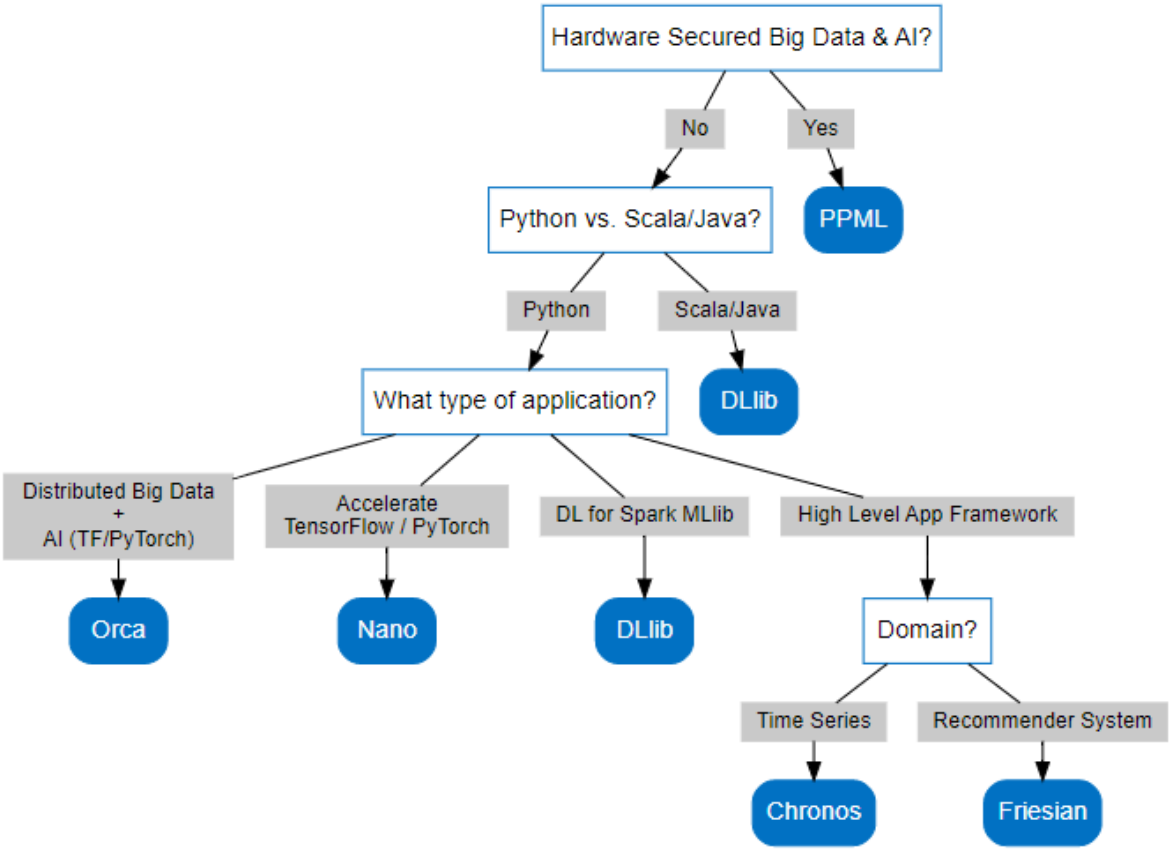
Bên dưới là bảng so sánh tốc độ running của các mô hình object detection.

Method	mAP	FPS	batch size	# Boxes	Input resolution
Faster R-CNN (VGG16)	73.2	7	1	~ 6000	~ 1000 × 600
Fast YOLO	52.7	155	1	98	448 × 448
YOLO (VGG16)	66.4	21	1	98	448 × 448
SSD300	74.3	46	1	8732	300 × 300
SSD512	76.8	19	1	24564	512 × 512
SSD300	74.3	59	8	8732	300 × 300
SSD512	76.8	22	8	24564	512 × 512

Hình 2.6: Bảng so sánh tốc độ xử lý và độ chính xác của các lớp model [20]

2.3 Mô hình mã nguồn mở BigDL

2.3.1 Tổng quan về BigDL



Hình 2.7: BigDL

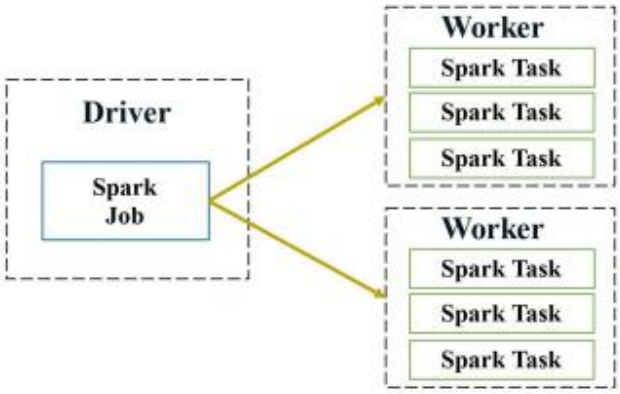
2.3.2 Mô hình thực thi BigDL

Trong khi sử dụng các phương pháp tiêu chuẩn như huấn luyện song song dữ liệu, máy chủ tham số để huấn luyện có khả năng mở rộng, điểm mới của BigDL là cách triển khai hiệu quả các chức năng này trên một mô hình tính toán cấu trúc của Apache Spark.

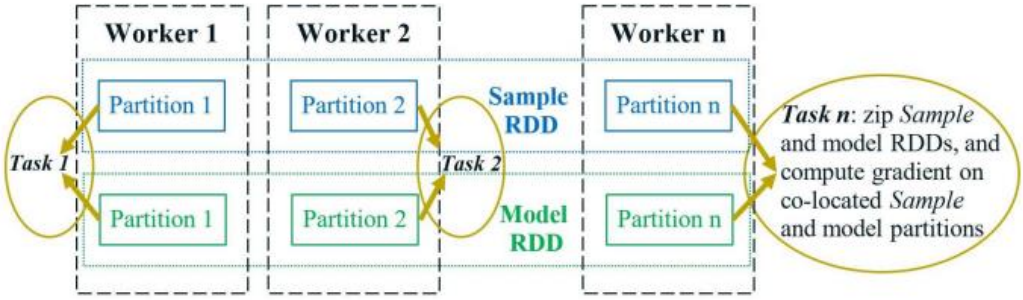
Trong cộng đồng học máy, truy cập dữ liệu chi tiết và thực hiện sửa đổi dữ liệu tại chỗ được xem là cực kỳ quan trọng để hỗ trợ cho việc huấn luyện phân tán hiệu quả với các máy chủ tham số. Tuy nhiên, trong hệ thống big data như Spark, mô hình tính toán cấu trúc khác được áp dụng, trong đó bộ dữ liệu không thay đổi và chỉ có thể được chuyển đổi thành bộ dữ liệu mới mà không có tác động phụ (tức là sao chép khi cần thiết); ngoài ra, các phép biến đổi đều là các thao tác cấu trúc thô (tức là áp dụng cùng

một phép biến đổi cho tất cả các mục dữ liệu cùng một lúc).

➢ ***Mô hình tính toán của Spark***



Hình 2.8: Mô hình của Spark: Driver Node có chức năng lập lịch và phân công công việc cho các Worker Node

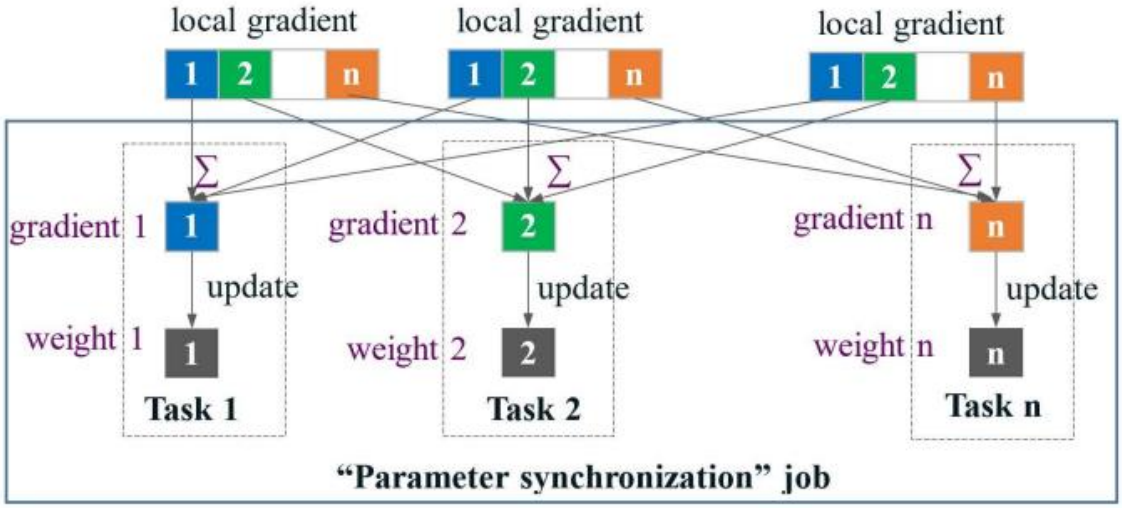


Hình 2.9: Tác vụ "forward-backword" của Spark tính toán gradient cho mỗi bản sao mô hình mạng nơ-ron song song

BigDL không hỗ trợ phân tán mô hình (model parallelism) tức là không có việc phân phối mô hình trên các worker khác nhau. Tuy nhiên, điều này không gây hạn chế trong thực tế, vì BigDL chạy trên các máy chủ Intel Xeon CPU, thường có dung lượng bộ nhớ lớn (100s GB) và có thể dễ dàng chứa các mô hình rất lớn.

➢ ***Đồng bộ hóa tham số trong BigDL***

Đồng bộ hóa tham số là một phép tính quan trọng đối với huấn luyện mô hình phân tán song song trên dữ liệu (về tốc độ và khả năng mở rộng). Để hỗ trợ đồng bộ hóa tham số hiệu quả, các framework học sâu hiện có thường triển khai máy chủ tham số hoặc AllReduce bằng cách sử dụng các phép tính như truy cập dữ liệu chi tiết và thay đổi dữ liệu tại chỗ. Thật không may, các phép tính này không được hỗ trợ bởi mô hình tính toán chức năng của hệ thống dữ liệu lớn (như Spark).



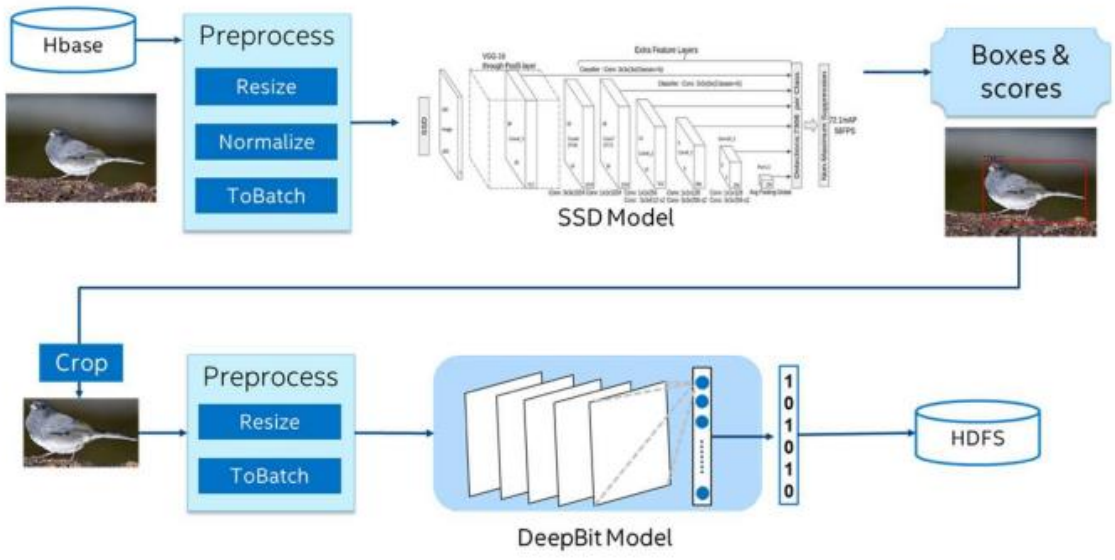
Hình 2.10: Đồng bộ hóa tham số trong BigDL

2.3.3 Ứng dụng BigDL cho bài toán nhận diện và phân loại hình ảnh

BigDL có thể được sử dụng để giải quyết bài toán nhận diện và phân loại hình ảnh bằng cách sử dụng các mô hình như SSD và DeepBit, được thể hiện trong hình 2.12.

Đầu tiên, sử dụng BigDL SSD model để phát hiện các vật thể trong hình ảnh. Các thông tin về địa điểm và kích thước của các vật thể này có thể được trích xuất và sử dụng để cắt ra các hình ảnh đầy đủ và tập trung vào các đối tượng cần nhận diện.

Sau đó, sử dụng BigDL DeepBit model để trích xuất các đặc trưng của từng hình ảnh đã được cắt ra từ bước trên. Các đặc trưng này có thể được sử dụng để phân loại ảnh thành các loại khác nhau và lưu trữ kết quả (RDD của các đặc trưng đối tượng được trích xuất) trong HDFS.



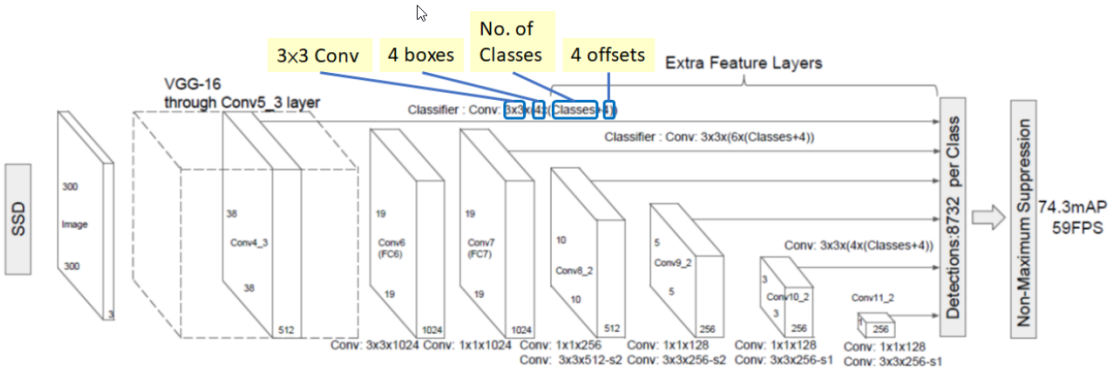
Hình 2.11: Ứng dụng BigDL với bài toán nhận diện và phân loại hình ảnh

❖ SSD Model

Mô hình SSD được xây dựng trên cơ sở của mạng neural tích chập (Convolutional Neural Network - CNN) và các lớp tích chập bổ sung. Kiến trúc của mô hình SSD bao gồm hai phần chính:

- Base Network: Một mạng neural tích chập (CNN) được sử dụng để xử lý ảnh đầu vào và trích xuất các đặc trưng của ảnh. Mạng CNN thường được huấn luyện trước trên một bộ dữ liệu lớn như ImageNet để trích xuất các đặc trưng có giá trị từ ảnh.
- MultiBox Head: Các lớp tích chập bổ sung được thêm vào sau mạng CNN để dự đoán các bounding box và xác suất lớp cho các đối tượng trong ảnh. MultiBox Head bao gồm các lớp tích chập và kết nối đầy đủ (fully connected) để biến đổi đầu vào từ mạng CNN thành các vector đặc trưng dùng để dự đoán vị trí và lớp của đối tượng. Bounding box được dự đoán bằng cách áp dụng một số lượng đặc trưng trên từng vị trí trên ảnh và dự đoán vị trí và kích thước của bounding box. Xác suất lớp cho các đối tượng được dự đoán bằng cách áp dụng một số lượng đặc trưng trên từng vị trí trên ảnh và tính xác suất đối tượng thuộc các lớp đã biết.

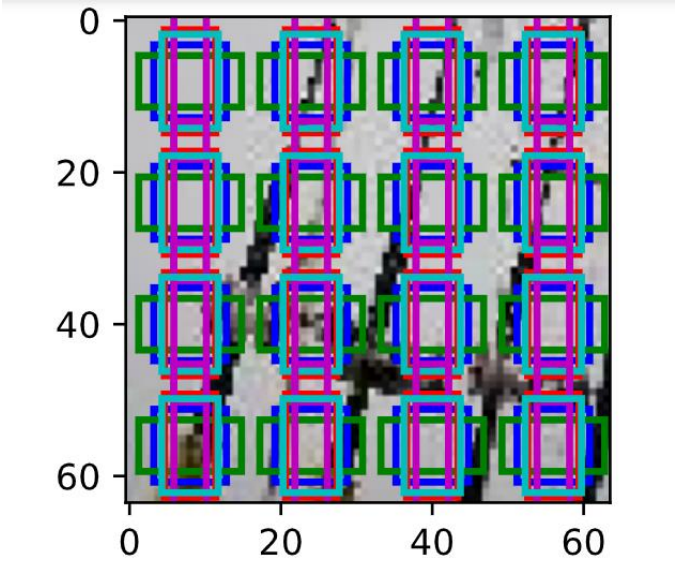
Với kiến trúc này, SSD có thể dự đoán bounding box và xác suất lớp tương ứng cho tất cả các đối tượng trong ảnh trong một lần chạy (single shot), giúp cho việc phát hiện đối tượng nhanh chóng và tiết kiệm tài nguyên tính toán.



Hình 2.12: Sơ đồ kiến trúc của mạng SSD [20]

SSD dựa trên việc áp dụng một kiến trúc chuẩn (Ví dụ: VGG16) để thực hiện

tiến trình lan truyền thuận và tạo ra một khối feature map 3D ở giai đoạn sớm. Kiến trúc mạng này được gọi là "base network" (từ input Image đến Conv7). Sau đó, chúng ta thêm các kiến trúc phía sau "base network" để tiến hành phát hiện vật thể, được gọi là "Extra Feature Layers" trong sơ đồ. Các lớp này có thể được giải thích một cách đơn giản như sau:



Hình 2.13: Vị trí của các default bounding box trên bức ảnh gốc khi áp dụng trên feature map có kích thước 4 x 4.

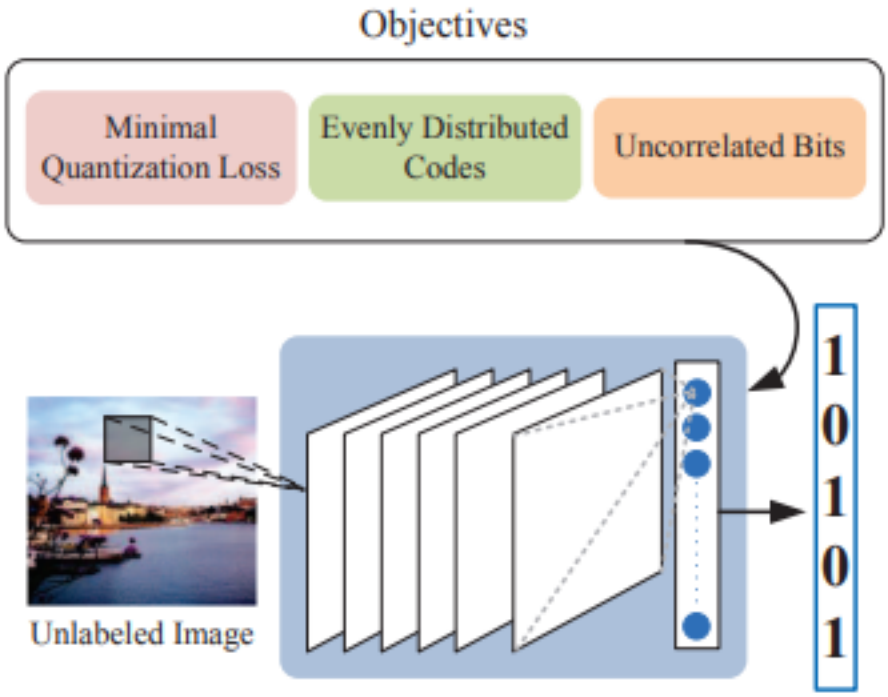
Như vậy, mỗi ô lưới trên feature map sẽ có kích thước là 4x4 và sẽ được liên kết với 4 default bounding box khác nhau như được minh họa trên hình vẽ. Tất cả các bounding box này có tâm trùng nhau và chính là tọa độ tâm của ô lưới mà chúng liên kết.

Tại mỗi một default bounding box trên feature map, chúng ta sẽ dự báo 4 offsets tương ứng với tọa độ và kích thước của nó. Các offsets này được biểu diễn bởi một tọa độ gồm 4 tham số (cx, cy, w, h), trong đó (cx, cy) xác định tọa độ tâm và (w, h) xác định kích thước của bounding box. Phần thứ hai trong dự báo là điểm số của bounding box tương ứng với mỗi lớp. Lưu ý rằng chúng ta sẽ có một lớp thứ C+1 để đại diện cho trường hợp mà default bounding box không chứa vật thể (hoặc thuộc lớp background).

Tương tự như anchor boxes trong mạng faster R-CNN, default boxes cũng được sử dụng trên một vài feature maps với các độ phân giải khác nhau. Điều này giúp cho các default bounding box có thể phân biệt hiệu quả kích thước của các vật thể khác

nhau.

❖ **DeepBit Model**



Hình 2.14: DeepBit Model

Mô hình xây dựng bộ giải mã bằng cách đặt cửa sổ chiếu lên ảnh đầu và và nhị phân hóa kết quả

$$b = 0.5 \times (\text{sign}(\mathcal{F}(x; \mathcal{W})) + 1), \qquad (1)$$

x đại diện cho ảnh đầu vào, b là bộ giải mã trong dạng vector. Sign(k)=1 nếu k>0 và bằng -1 nếu ngược lại. F(x, W) là 1 tập hợp các chức năng chiếu xuống có thể viết như sau:

$$\mathcal{F}(x; \mathcal{W}) = f_k(\cdots f_2(f_1(x; w_1); w_2) \cdots ; w_k), \qquad (2)$$

f lấy dữ liệu x_i và tham số w_i là đầu vào và tạo ra kết quả chiếu xuống x_{i+1}

Cách giải quyết này dùng giúp thông số đối chiếu W=(w1,w2,w3,…wn) lượng tử hóa hình ảnh đầu vào x thành 1 vector b nhị phân gọn nhẹ mà không làm mất thông tin từ đầu vào. Để tạo 1 bộ giải mã gọn nhẹ và phân biệt tốt, bộ giải mã phải có sự mất mát lượng tử hóa ít nhất để giữ được cấu trúc dữ liệu từ lớp trước. Thứ hai, bộ giải mã phải phân bố đồng đều để xâu nhị phân phân biệt được nhiều thông điệp khác nhau

hơn. Cuối cùng bộ giải mã phải bất biến trước sự xoay hay nhiễu của vật thể, từ đó bộ giải mã có thể bắt được nhiều thông tin hơn từ mội ảnh.

2.4 Kết luận chương

Tại chương này, luận văn đã trình bày tổng quan các phương pháp nhận diện đặt ra và lý thuyết cho hệ thống.

Đầu tiên, luận văn đã giải thích các khái niệm cơ bản về nhận diện đối tượng và học sâu, cùng với kiến trúc mạng phổ biến như Convolutional Neural Networks (CNNs).

Tiếp theo, luận văn cũng đã giới thiệu về BigDL - một thư viện học sâu mã nguồn mở - và cách sử dụng BigDL cho bài toán nhận diện hình ảnh. Sử dụng BigDL có thể tăng tốc độ đào tạo và chạy mô hình, đồng thời cũng giảm thiểu thời gian xử lý dữ liệu. Từ những kiến thức và kinh nghiệm đã tìm hiểu được trong chương này, luận văn sẽ ứng dụng BigDL cho bài toán nhận diện và phân loại nông sản ở chương 3.

CHƯƠNG 3 . KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ

3.1 Thu thập dữ liệu

Cơ sở bao gồm các ảnh chụp nông sản ở các góc độ khác nhau với nền tùy ý, được lấy từ các dataset trên mạng hoặc tự chụp bằng thiết bị camera cá nhân.

Danh sách các nông sản bao gồm:

'Táo', 'Mơ', 'Chuối', 'Cải Bắp', 'Dưa lưới', 'Khế', 'Cà rốt', 'Súp lơ', 'Dừa', 'Ngô', 'Cà tím', 'Tỏi', 'Gừng', 'Nho', 'Bưởi', 'Ổi', 'Su hào', 'Quất', 'Chanh', 'Nhãn', 'Vải', 'Xoài', 'Măng cụt', 'Dâu tằm', 'Hành tây', 'Cam', 'Đào', 'Lê', 'Ớt chuông ', 'Hồng', 'Dứa', "Thanh long', 'Mận', 'Lựu', 'Khoai tây', 'Chôm chôm', 'Hồng xiêm', 'Dâu tây', 'Cà chua', 'Dưa hấu',

3.2 Thực nghiệm với các phương pháp

- **Thực nghiệm với phương pháp Học máy truyền thống**
  - **Bước 1. Chuẩn bị dữ liệu:** Xây dựng CSDL ảnh nông sản cho 20 loại nông sản, kèm theo các nhãn cho từng hình ảnh để đánh dấu loại nông sản tương ứng
  - **Bước 2**. **Tiền xử lý ảnh:** Ứng dụng mô hình SSD Model chuẩn bị dữ liệu trước khi đưa vào mô hình để huấn luyện hoặc dự đoán.
  - **Bước 3. Chọn lọc đặc trưng cụ thể:** Mỗi ảnh đầu vào ta sẽ tính toán được 30 giá trị đại diện cho 30 đặc trưng về màu sắc, hình dạng và kết cấu. Những đặc trưng này được chọn lựa sau quá trình tìm hiểu các bài báo, công trình khoa học về sử dụng Học máy trong bài toán nhận dạng nông sản và thống kê các đặc trưng được sử dụng nhiều nhất, đạt hiệu quả tốt nhất. [2][3][4]
  - **Bước 4:** Huấn luyện mô hình nhận dạng nông sản từ CSDL ảnh đã xây dựng. Bộ CSDL ảnh này chỉ để so sánh tương đối độ chính xác của mô hình truyền thống so với mô hình học sâu tiên tiến bây giờ, do đó số lượng loại nông sản được hạn chế chỉ còn 20 loại, với số lượng ảnh cho mỗi loại là 400-600 ảnh.
  - **Bước 5:** Thống kê độ chính xác của bộ test với tỉ lệ bộ training/test là 75/25.
- **Thực nghiệm với phương pháp Học sâu (sử dụng BIGDL)**:
  - **Bước 1. Chuẩn bị dữ liệu:** Xây dựng CSDL ảnh nông sản cho 20 loại nông sản, kèm theo các nhãn cho từng hình ảnh để đánh dấu loại nông sản tương ứng

- **Bước 2. Tiền xử lý ảnh:** Ứng dụng mô hình SSD Model chuẩn bị dữ liệu trước khi đưa vào mô hình để huấn luyện hoặc dự đoán.

- **Bước 3. Xây dựng mô hình:** Sử dụng BigDL để xây dựng mô hình phân loại nông sản. Mô hình sử dụng SSD model để phát hiện vật thể trong ảnh, sau đó sử dụng DeepBit model để rút trích đặc trưng và phân loại nông sản.

- **Bước 4. Huấn luyện mô hình:** Sử dụng tập dữ liệu đã chuẩn bị để huấn luyện mô hình. Sử dụng thuật toán Stochastic Gradient Descent (SGD) để tối ưu hóa hàm loss function.

- **Bước 5. Đánh giá mô hình:** Thống kê độ chính xác của bộ test với tỉ lệ bộ training/test là 75/25.

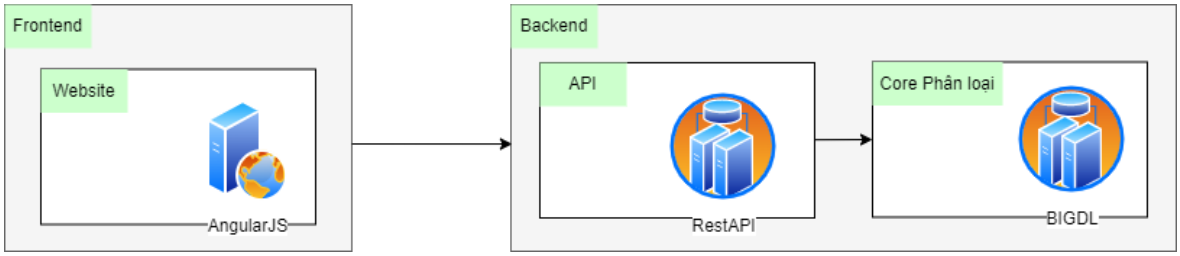
Thông tin tổng quan về bộ CSDL ảnh và quá trình huấn luyện cũng như kết quả đạt được của hai phương pháp cũng được tóm lược trong bảng bên dưới:

Bảng 3.1: So sánh sơ bộ kết quả huấn luyện của 2 phương pháp

	Thời gian huấn luyện	**Độ chính xác**
Học máy truyền thống	~ 30 phút	71,54%
Học sâu (sử dụng BIGDL)	~ 60 phút	~94.61%

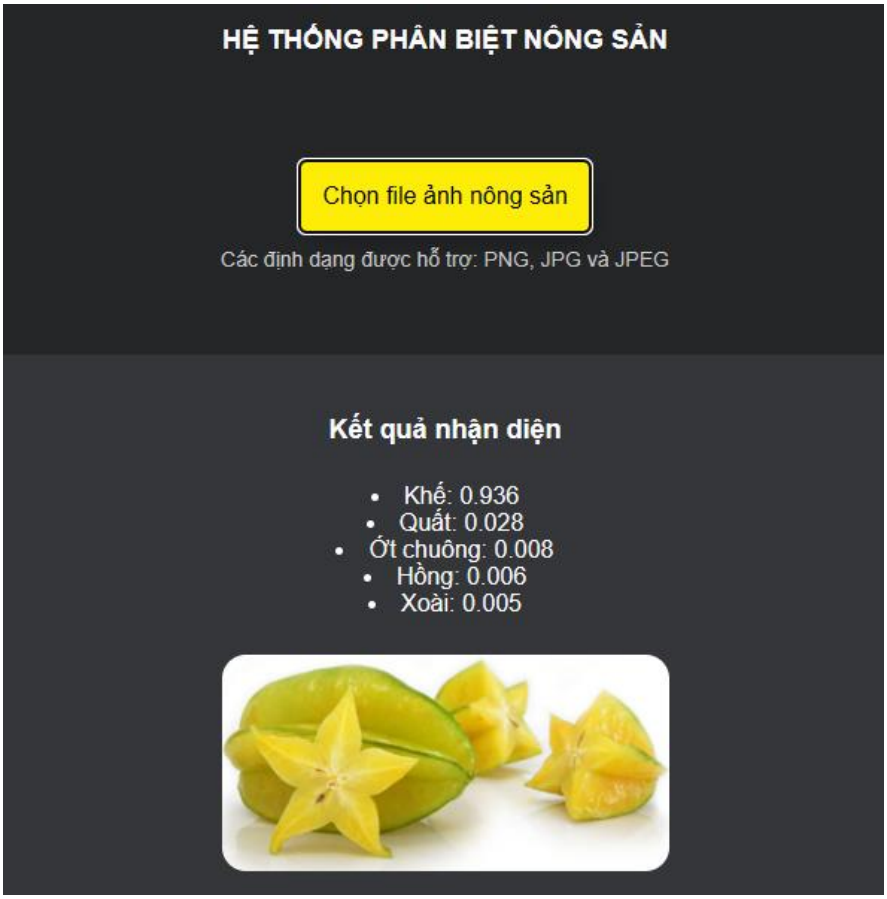
3.3 Ứng dụng nhận diện và phân loại nông sản

Ứng dụng nhận diện và phân loại nông sản được xây dựng theo mô hình Frontend/Backend bao gồm 2 thành phần chính:



Hình 3.2: Mô hình Ứng dụng Nhận dạng nông sản

3.4 Kết quả



Hình 3.3: Kết quả nhận dạng tốt với loại nông sản có đặc trưng riêng biệt

Đối với những loại nông sản có nhiều nét tương đồng lẫn nhau, kết quả nhận dạng của ứng dụng còn đôi lúc bị nhầm lẫn, đặc biệt trong các trường hợp ảnh được chụp theo góc nhìn chưa tốt dẫn đến ảnh không thể hiện được các đặc trưng riêng của quả. Nguyên nhân dẫn đến nhầm lẫn bao gồm:

- *Thiếu dữ liệu đa dạng:* Nếu tập dữ liệu sử dụng để huấn luyện mô hình không đủ đa dạng, mô hình có thể không nhận diện được các đặc điểm khác biệt giữa các loại nông sản.
- *Sai số trong quá trình thu thập dữ liệu:* Nếu dữ liệu được sử dụng để huấn luyện mô hình không chính xác hoặc không đầy đủ, các đặc trưng quan trọng có thể bị bỏ qua và dẫn đến sự nhầm lẫn.
- *Độ phân giải ảnh không đủ:* Nếu độ phân giải của ảnh được sử dụng để huấn luyện mô hình không đủ, các chi tiết nhỏ hoặc đặc trưng quan trọng có thể bị mất đi, dẫn đến sự nhầm lẫn.

- *Địa hình và điều kiện ánh sáng:* Nếu hình ảnh được chụp trong điều kiện ánh sáng kém hoặc trên địa hình khác nhau, các đặc trưng quan trọng có thể không được nhận diện, dẫn đến sự nhầm lẫn.
- *Kiến trúc mô hình không phù hợp:* Nếu kiến trúc mô hình được sử dụng không phù hợp với bài toán hoặc tập dữ liệu cụ thể, mô hình có thể không đưa ra các kết quả chính xác.

Hình 3.4: Kết quả nhận dạng chưa tốt với loại quả không có đặc trưng riêng biệt

Hình 3.5: Kết quả nhận dạng với loại quả không được huấn luyện

Trong trường hợp như hình trên, khi yêu cầu hệ thống nhận dạng quả bòn bon, do bòn bon không có trong danh sách nông sản được huấn luyện nhận dạng nên kết quả trả về là loại quả có sự tương đồng cao nhất, quả nhãn.

Ngoài ra, kết quả thực nghiệm thu được cho thấy hệ thống nhận dạng đạt được kết quả tương đối chuẩn xác với các trường hợp hình ảnh quả trong ảnh đầu vào bị che khuất một phần, điều kiện ánh sáng không thực sự tốt cũng như các trường hợp ảnh bị biến dạng nhẹ. Đây chính là các khó khăn đối với bài toán nhận dạng vật thể nói chung mà luận văn đã đề cập tới trong phần mở đầu, lý giải cho điều này là do trong quá trình thu thập ảnh ban đầu cũng như sinh ảnh tự động từ các ảnh gốc, mô hình nhận dạng đã được huấn luyện để nhận ra các trường hợp tương tự. Khả năng dự đoán mạnh mẽ này đã giúp cho các phương pháp Học sâu, đặc biệt là mạng huấn luyện nơron tích chập SSD trở thành giải pháp mạnh mẽ nhất trong lĩnh vực nhận dạng ảnh bây giờ.

3.5 Kết luận chương

Kết thúc chương, luận văn đã nghiên cứu, tìm hiểu bài toán tự động nhận dạng và phân loại nông sản trong ảnh màu, và thực hiện phát triển, cài đặt phương án giải quyết cho bài toán dựa trên sự thống kê các hướng tiếp cận đã được công bố qua rất

nhiều bài báo, công trình khoa học trên thế giới. Các kết quả chính mà luận văn đã đạt được, tương ứng với các mục tiêu đề ra.

KẾT LUẬN

Đề tài luận văn đã nghiên cứu về mô hình học sâu như CNN cũng như tiến hành xây dựng cơ sở dữ liệu ảnh nông sản và phát triển một hệ thống nhận diện nông sản sử dụng BigDL. Qua quá trình nghiên cứu và thực hiện, luận văn đã đạt được những kết quả đáng kể và đối mặt với một số khó khăn.

Việc xây dựng cơ sở dữ liệu ảnh nông sản là một bước quan trọng, đảm bảo nguồn dữ liệu phong phú và đa dạng để huấn luyện và đánh giá mô hình nhận diện. Luận văn đã thu thập ảnh từ các dataset có sẵn và tự chụp ảnh bằng thiết bị cá nhân, đồng thời tổ chức dữ liệu theo các thông tin nhãn, định danh và vị trí chụp để tạo ra cơ sở dữ liệu có tổ chức.

Trong quá trình nghiên cứu, luận văn đã tiếp cận và thử nghiệm các phương pháp trí tuệ nhân tạo (bao gồm cả học máy và học sâu) để nhận diện nông sản. Bằng việc sử dụng mô hình BIGDL, hệ thống nhận diện nông sản đã được xây dựng và kiểm thử, và kết quả đạt được đã chứng minh hiệu quả và độ chính xác của phương pháp đề xuất.

Tuy nhiên, luận văn cũng đã đối mặt với một số khó khăn trong quá trình nghiên cứu và thực hiện. Việc thu thập dữ liệu ảnh nông sản có thể mất nhiều thời gian và công sức, đồng thời gán nhãn và xử lý dữ liệu cũng đòi hỏi sự tỉ mỉ và kiên nhẫn. Ngoài ra, việc tinh chỉnh mô hình và xử lý ảnh không đồng nhất cũng là những thách thức mà luận văn đã phải đối mặt.

Tổng hợp lại, đề tài luận văn đã đạt được những kết quả quan trọng trong việc xây dựng cơ sở dữ liệu ảnh nông sản, nghiên cứu các đặc trưng và xây dựng mô hình nhận diện. Các kết quả này mang lại sự tiện ích và ứng dụng trong việc phân loại và nhận diện nông sản, góp phần nâng cao hiệu quả và tự động hóa trong lĩnh vực nông nghiệp.