

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

NGUYỄN ĐẮC THÀNH

NHẬN DẠNG VÀ PHÂN LOẠI HOA QUẢ TRONG ẢNH MÀU

Ngành: Công nghệ thông tin

Chuyên ngành: Kỹ thuật phần mềm

Mã số: 60480103

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT PHẦN MỀM

Hà Nội – 2016

Mục lục

MỞ ĐẦU	3
1. Tính cấp thiết của đề tài luận văn	3
2. Mục tiêu của luận văn	3
2.1. Cơ sở dữ liệu ảnh hoa quả	3
2.2. Bộ huấn luyện nhận dạng hoa quả	4
2.3. Ứng dụng nhận dạng hoa quả	5
3. Cấu trúc của luận văn	5
Chương 1. Giới thiệu tổng quan	7
1.1. Bài toán nhận dạng và phân loại hoa quả	7
1.2. Các hướng tiếp cận và giải quyết bài toán	7
1.2.1. Phương pháp Machine Learning truyền thống	8
1.2.2. Phương pháp Học sâu	10
Chương 2. Mạng nơ-ron tích chập	11
2.1. Kiến trúc mạng nơ-ron tích chập	11
2.2. Học chuyển giao và tinh chỉnh mô hình huấn luyện	13
2.3. Mạng huấn luyện AlexNet	14
2.2.1. Kiến trúc mạng AlexNet	15
2.4. Ứng dụng mạng AlexNet vào bài toán Nhận dạng, phân loại hoa quả	16
Chương 3. Tổng quan Hệ thống Fruit Recognition System	17
3.1. Tổng quan hệ thống	17

3.2. Mô đun quản lý cơ sở dữ liệu.....	18
3.3. Bộ huấn luyện mô hình	18
3.3.1. Môi trường huấn luyện	19
3.3.2. Cấu hình mạng huấn luyện AlexNet.....	20
3.3.3. Một số hình ảnh về đặc trưng do mạng AlexNet tính toán	20
3.3.4. Các mô đun phía Server	21
3.4. Ứng dụng phía Client	22
Chương 4. Kết quả thử nghiệm và đánh giá	23
4.1. So sánh với phương pháp Machine Learning truyền thống	23
4.2. So sánh kết quả với bộ CSDL được sinh tự động	24
4.3. Thử nghiệm ứng dụng trong thực tế.....	26
Chương 5. Kết luận.....	28

MỞ ĐẦU

1. Tính cấp thiết của đề tài luận văn

Hiện nay, ở nước ta nói riêng và ở các nước đang phát triển có nền nông nghiệp là một trong các ngành sản xuất chủ yếu, quá trình thu hoạch, phân loại và đánh giá chất lượng các loại sản phẩm nông nghiệp, đặc biệt là các loại hoa quả, chủ yếu còn phải thực hiện bằng các phương pháp thủ công. Đây là công việc không quá khó, nhưng tiêu tốn nhiều thời gian, công sức của con người và là rào cản đối với mở rộng phát triển quy mô sản xuất nông nghiệp.

Trong thời gian gần đây, nhờ có sự phát triển mạnh mẽ về khả năng tính toán của các thế hệ máy tính hiện đại cũng như sự bùng nổ về dữ liệu thông qua mạng lưới Internet trải rộng, ta đã chứng kiến nhiều sự đột phá trong lĩnh vực Học máy, đặc biệt là trong lĩnh vực Thị giác máy tính. Sự quay lại và phát triển vượt bậc của các phương pháp Học sâu đã giúp Thị giác máy tính đạt được những thành tựu đáng kể trong lĩnh vực Nhận dạng ảnh, trong đó có bài toán nhận dạng hoa quả. Đề tài nghiên cứu “Nhận dạng và phân loại hoa quả trong ảnh màu” 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 dạng hoa quả tự động, đặc biệt là đối với các loại hoa quả phổ biến tại nước ta.

2. Mục tiêu của luận văn

2.1. Cơ sở dữ liệu ảnh hoa quả

Bộ cơ sở dữ liệu ảnh là một trong các thành phần quan trọng hàng đầu trong các phương pháp Học máy nói chung, được sử dụng

để phục vụ cho quá trình tính toán tham số và huấn luyện, tinh chỉnh các mô hình. Thông thường, bộ dữ liệu càng lớn và càng được chọn lọc tỉ mỉ cẩn thận thì độ chính xác của mô hình càng được cải thiện, nhưng trong phạm vi luận văn này sẽ được hạn chế cả về số lượng loại hoa quả sẽ nhận dạng cũng như số lượng ảnh chụp cho mỗi loại hoa quả đó. Cụ thể:

- Số lượng hoa quả sẽ nhận dạng: 40 loại hoa quả phổ biến ở nước ta như nho, táo, chuối, thanh long...
- Số lượng ảnh gốc cho mỗi loại quả: 500-1000 ảnh

Sau khi đã thu thập đủ số lượng ảnh gốc cho các loại hoa quả, ta sẽ sử dụng các thuật toán chỉnh sửa ảnh, như làm nghiêng ảnh, chèn thêm nhiễu hoặc ghép ảnh với nền khác, để tạo thêm ảnh mới nhằm tăng cường kích thước cơ sở dữ liệu (CSDL).

2.2. Bộ huấn luyện nhận dạng hoa quả

Để đưa ra đánh giá tổng quát và so sánh độ chính xác tương đối giữa các phương pháp Học máy truyền thống với phương pháp Học sâu, luận văn thực hiện cài đặt một mạng huấn luyện nơ-ron nhân tạo truyền thống và một mạng huấn luyện nơ-ron tích chập trong Học sâu, sau khi thực hiện huấn luyện trên cùng bộ cơ sở dữ liệu ảnh và so sánh kết quả.

Đối với phương pháp Học máy truyền thống: nghiên cứu, tìm hiểu các phương pháp đã được trình bày trong các bài báo, công trình khoa học và thống kê ra các đặc trưng thường được sử dụng và cho kết quả huấn luyện tốt nhất. Các đặc trưng này thể hiện thông tin của

hoa quả về màu sắc, hình dạng và kết cấu, và được đưa vào bộ tính toán, trích chọn đặc trưng của mạng nơ-ron nhân tạo.

Đối với mạng nơ-ron tích chập thuộc nhóm Học sâu: tìm hiểu và chọn một trong các mô hình huấn luyện phổ biến trong lĩnh vực Nhận dạng ảnh trên thế giới để thực hiện cài đặt và so sánh kết quả với bộ nhận dạng truyền thống.

2.3. Ứng dụng nhận dạng hoa quả

Một trong các mục tiêu của luận văn là xây dựng thành công một ứng dụng đơn giản trên smartphone nhằm hỗ trợ người dùng nhận dạng hoa quả. Hệ thống nhận dạng hoa quả - Fruit Recognition System - ngoài ứng dụng client trên smartphone còn có một máy chủ server để thực hiện tất cả các bước huấn luyện và nạp mô hình nhận dạng, các bước tính toán nhận dạng loại hoa quả dựa trên ảnh chụp nhận được từ ứng dụng client.

3. Cấu trúc của luận văn

Dựa trên mục tiêu cụ thể đã trình bày trong phần trước, luận văn được tổ chức thành năm chương với các nội dung cụ thể như sau:

- Chương 1: Trong chương tổng quan này, ta sẽ có ra cái nhìn tổng quan về các hướng tiếp cận và giải pháp đã được ứng dụng trong bài toán nhận dạng phân loại hoa quả, từ các phương pháp thuần tính toán xử lý ảnh tương đối thô sơ cho tới các phương pháp Học máy truyền thống và cuối cùng là các phương pháp Học sâu.

- Chương 2: Chương này sẽ đi sâu hơn vào một mạng huấn luyện trong Học sâu thường được sử dụng trong lĩnh vực Nhận dạng

ảnh - mạng nơ-ron tích chập, và tìm hiểu chìa khóa giải quyết bài toán nhận dạng ảnh với bộ dữ liệu huấn luyện có kích thước tương đối nhỏ.

- Chương 3: Trong chương tiếp theo, ta sẽ mô tả tổng quan Hệ thống nhận dạng hoa quả tự động do nhóm đề tài phát triển, với các mô đun chính như máy chủ, máy trạm, bộ huấn luyện và nhận dạng, bộ tiền xử lý... Ngoài ra, cách thức thu thập, chỉnh sửa cơ sở dữ liệu ảnh và cách cài đặt, triển khai môi trường huấn luyện cho mô hình mạng nơ-ron tích chập đã chọn trong chương 2 cũng sẽ được trình bày cụ thể tại đây.

- Chương 4: Chương 4 tập trung trình bày về kết quả thực nghiệm, bao gồm kết quả so sánh độ chính xác giữa các phương pháp Học máy truyền thống với phương pháp Học sâu, cùng với các đánh giá về độ hiệu quả của bộ tạo dữ liệu ảnh nhiễu cũng như các ảnh chụp ứng dụng khi được sử dụng trong thực tế. Dựa trên các kết quả thực nghiệm này, ta sẽ đưa ra một số phân tích và kết luận về điểm mạnh và điểm hạn chế của mô hình huấn luyện Học sâu đã chọn.

- Chương 5: Cuối cùng, chương 5 sẽ tổng kết các nội dung đã trình bày trong luận văn, từ đó đề xuất các phương hướng nghiên cứu tiếp theo để tiếp tục cải thiện chất lượng nhận dạng của hệ thống.

Chương 1. Giới thiệu tổng quan

1.1. Bài toán nhận dạng và phân loại hoa quả

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. Những khó khăn tiêu biểu có thể kể đến: sự đa dạng điểm nhìn, sự đa dạng kích thước, điều kiện ánh sáng khác biệt, sự ẩn giấu vật thể sau các đối tượng khác trong ảnh, sự lẫn lộn với nền và sự đa dạng về chủng loại vật thể...

Là một trường hợp cụ thể của bài toán nhận dạng và phân lớp, bài toán nhận dạng hoa quả kế thừa các khó khăn vốn có của bài toán gốc, và kèm theo là các khó khăn riêng của chính nó, như: số lượng khổng lồ về chủng loại hoa quả theo mùa, vùng miền, địa hình... với vô số loại hoa quả có hình dáng, màu sắc, kết cấu giống nhau, dải biến thiên màu sắc theo chu kỳ phát triển của quả từ lúc còn xanh đến lúc chín, hay sự đa dạng về hình dạng của cùng một loại quả do ảnh hưởng của thời tiết, điều kiện thổ nhưỡng và chế độ dinh dưỡng...

1.2. Các hướng tiếp cận và giải quyết bài toán

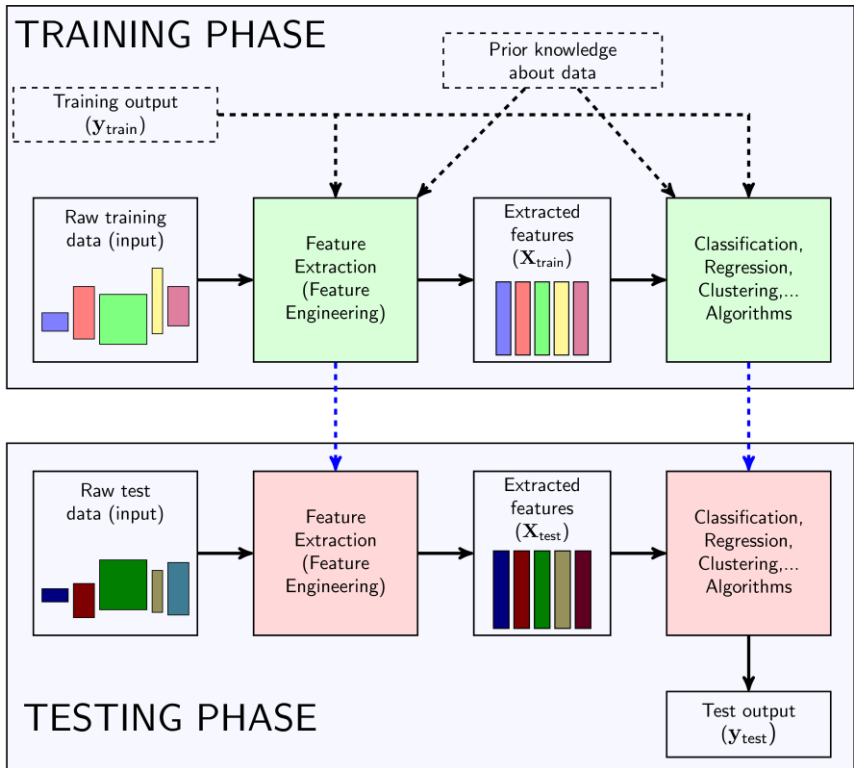
Bài toán tự động nhận dạng hoa quả đã xuất hiện từ lâu và đã có rất nhiều bài báo, công trình khoa học được đưa ra nhằm đề xuất

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 hoa quả. 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 hoa quả quá phức tạp, kết quả đạt được của các phương pháp này không được cao và phạm vi áp dụng trên số lượng loại hoa quả cũng bị hạn chế.

Bắt đầu từ những năm 2000s, sau khi xuất hiện một vài bài báo khoa học đề xuất áp dụng phương pháp Học máy - Machine Learning - vào bài toán nhận dạng hoa quả với độ chính xác cao, hướng giải quyết bài toán đã tập trung vào ứng dụng và cải tiến các thuật toán Học máy, cụ thể là nghiên cứu, thử nghiệm trích chọn các đặc trưng phù hợp nhất để đưa vào huấn luyện bộ nhận dạng tự động. Trong những năm gần đây, nhờ sự phát triển vượt bậc về sức mạnh tính toán của các máy tính cũng như sự bùng nổ dữ liệu trên Internet, Học sâu - Deep Learning đã đạt được nhiều thành tựu đáng kể, đặc biệt là trong lĩnh vực Xử lý ảnh và ngôn ngữ tự nhiên. Học sâu cũng đã được áp dụng rất thành công vào bài toán nhận dạng hoa quả, trong phạm vi hạn chế về số lượng loại hoa quả cần nhận dạng, phương pháp này đã đạt được kết quả cao đáng kinh ngạc.

1.2.1. Phương pháp Machine Learning truyền thống

Mô hình hoạt động chung của các phương pháp Machine Learning được thể hiện trong hình dưới đây:



Ta có thể thấy Machine Learning gồm hai giai đoạn chính là Huấn luyện – Training và Thử nghiệm – Testing, trong mỗi giai đoạn đều sử dụng hai thành phần quan trọng nhất do người xử lý bài toán thiết kế, đó là Trích chọn đặc trưng – Feature Engineering (hay còn gọi là Feature Extraction) và Thuật toán phân loại, nhận dạng... - Algorithms. Hai thành phần này có ảnh hưởng trực tiếp đến kết quả bài toán, vì thế được thiết kế rất cẩn thận, tốn nhiều thời gian, đòi hỏi người thiết kế phải có kiến thức chuyên môn và nắm rõ đặc điểm của bài toán cần xử lý.

1.2.1.1. Trích chọn đặc trưng - Feature Engineering

Trong các bài toán thực tế, ta chỉ có được những dữ liệu thô chưa qua chọn lọc xử lý, và để có thể đưa các dữ liệu này vào huấn luyện ta cần có những phép biến đổi để biến các dữ liệu thô thành dữ liệu chuẩn, không còn nhiễu và có khả năng biểu diễn dữ liệu tốt hơn. Các thông tin đặc trưng này là khác nhau với từng loại dữ liệu và bài toán cụ thể, vì thế trong từng trường hợp phép biến đổi này cần phải được tùy biến một cách thích hợp để cải thiện độ chính xác của mô hình dự đoán. Quá trình này được gọi là *Trích chọn đặc trưng*

1.2.2. Phương pháp Học sâu

Là 1 lĩnh vực chuyên sâu của Học máy, đã xuất hiện từ những năm 1980s nhưng chưa phổ biến cho đến thập kỷ gần đây do các nhà khoa học đã có thể tận dụng khả năng tính toán mạnh mẽ của các máy tính hiện đại cũng như khối lượng dữ liệu khổng lồ (hình ảnh, âm thanh, văn bản,...) trên Internet.

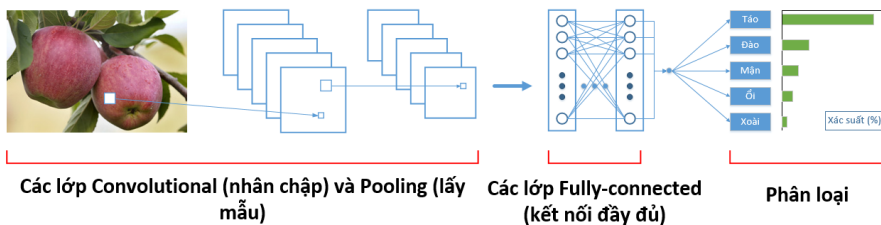
Các mạng huấn luyện theo phương pháp Học sâu còn được gọi với cái tên khác là mạng nơ-ron sâu (Deep Neural Network) do cách thức hoạt động của chúng. Về cơ bản, các mạng này bao gồm rất nhiều lớp khác nhau, mỗi lớp sẽ phân tích dữ liệu đầu vào theo các khía cạnh khác nhau và theo mức độ trừu tượng nâng cao dần.

Chương 2. Mạng nơ-ron tích chập

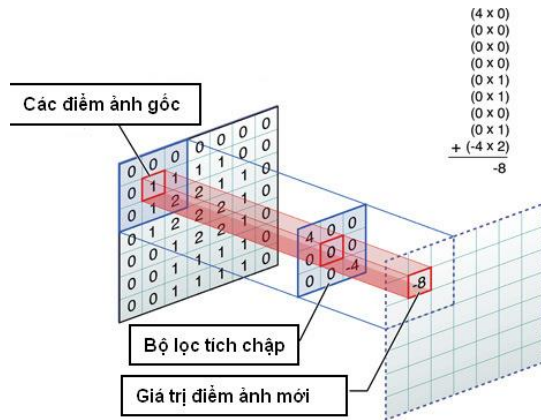
Mạng nơ-ron tích chập (CNN - Convolutional Neural Network) là một trong những mô hình mạng Học sâu phổ biến nhất hiện nay, có khả năng nhận dạng và phân loại hình ảnh với độ chính xác rất cao, thậm chí còn tốt hơn con người trong nhiều trường hợp. Mô hình này đã và đang được phát triển, ứng dụng vào các hệ thống xử lý ảnh lớn của Facebook, Google hay Amazon... cho các mục đích khác nhau như các thuật toán tagging tự động, tìm kiếm ảnh hoặc gợi ý sản phẩm cho người tiêu dùng.

2.1. Kiến trúc mạng nơ-ron tích chập

Các lớp cơ bản trong một mạng CNN bao gồm: Lớp Convolutional, Lớp ReLU (Rectified Linear Unit), Lớp Pooling và Lớp Fully-connected, được thay đổi về số lượng và cách sắp xếp để tạo ra các mô hình huấn luyện phù hợp cho từng bài toán khác nhau.



- **Lớp Convolutional** (tích chập): Đây là thành phần quan trọng nhất trong mạng CNN, cũng là nơi thể hiện tư tưởng xây dựng sự liên kết cục bộ thay vì kết nối toàn bộ các điểm ảnh. Các liên kết cục bộ này được tính toán bằng phép tích chập giữa các giá trị điểm ảnh trong một vùng ảnh cục bộ với các bộ lọc – filters – có kích thước nhỏ.



Trong ví dụ, ta thấy 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.

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 ReLU:** Lớp này được xây dựng với ý nghĩa đảm bảo tính phi tuyến của mô hình huấn luyện sau khi đã thực hiện một loạt các phép tính toán tuyến tính qua các lớp Tích chập. Trong số các hàm kích hoạt phổ biến nhất như tanh, sigmoid..., hàm ReLU được chọn do cài đặt đơn giản, tốc độ xử lý nhanh mà vẫn đảm bảo được tính toán hiệu quả. Cụ thể, phép tính toán của hàm ReLU chỉ đơn giản là chuyển tất cả các giá trị âm thành giá trị 0.

Thông thường, lớp ReLU được áp dụng ngay phía sau lớp Tích chập, với đầu ra là một ảnh mới có kích thước giống với ảnh đầu vào, các giá trị điểm ảnh cũng hoàn toàn tương tự trừ các giá trị âm đã bị loại bỏ.

- **Lớp Pooling:** Một thành phần tính toán chính khác trong mạng CNN là pooling, thường được đặt sau lớp Tích chập và lớp ReLU để làm giảm kích thước kích thước ảnh đầu ra trong khi vẫn giữ được các thông tin quan trọng của ảnh đầu vào. Có hai phương pháp lấy mẫu phổ biến hiện nay là lấy mẫu giá trị lớn nhất (Max Pooling) và lấy mẫu giá trị trung bình (Average Pooling).

- **Lớp Fully-connected:** Lớp Fully-connected này 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.

2.2. Học chuyên giao và tinh chỉnh mô hình huấn luyện

Trong thời gian đầu khi các phương pháp Học sâu mới đạt được nhiều thành tựu và được áp dụng phổ biến, trong cộng đồng Học sâu đã xuất hiện một quan niệm sai lầm: nếu bạn không có lượng dữ liệu huấn luyện khổng lồ, bạn không thể tạo ra một mô hình dự đoán hiệu quả. Học chuyên giao – Transfer Learning – chính là từ khóa để giải quyết điểm hạn chế này.

Học chuyên giao là quá trình khai thác, tái sử dụng các tri thức đã được học tập bởi một mô hình huấn luyện trước đó vào giải quyết một bài toán mới mà không phải xây dựng một mô hình huấn luyện khác từ đầu.

Hiện nay, phương pháp phổ biến thường được áp dụng với trường hợp bộ CSDL tương đối nhỏ là tận dụng một mạng CNN đã được huấn luyện trước đó với bộ dữ liệu rất lớn như ImageNet

- Mạng CNN này sẽ chỉ được sử dụng như một bộ trích chọn đặc trưng cho bộ CSDL huấn luyện mới, bằng cách thay thế các lớp Fully-connected ở cuối mạng và giữ cố định các tham số cho toàn bộ các lớp còn lại của mạng.

- Không chỉ thay thế và huấn luyện lại bộ nhận dạng cuối cùng của mạng CNN, mà đồng thời ta thực hiện tối ưu, tinh chỉnh (Fine-tune) một vài hoặc tất cả các lớp trong mạng.

Ý tưởng của việc tái sử dụng mạng CNN là dựa trên nhận định rằng các đặc trưng được học trong các lớp đầu của mạng là các đặc trưng chung nhất, hữu dụng với phần lớn bài toán, ví dụ: đặc trưng về cạnh, hình khối hay các khối màu... Các lớp sau đó của mạng CNN sẽ nâng dần độ cụ thể, riêng biệt của các chi tiết phục vụ cho bài toán nhận dạng ban đầu. Do đó, ta hoàn toàn có thể tái sử dụng lại các lớp đầu của mạng CNN mà không phải mất nhiều thời gian và công sức huấn luyện từ đầu.

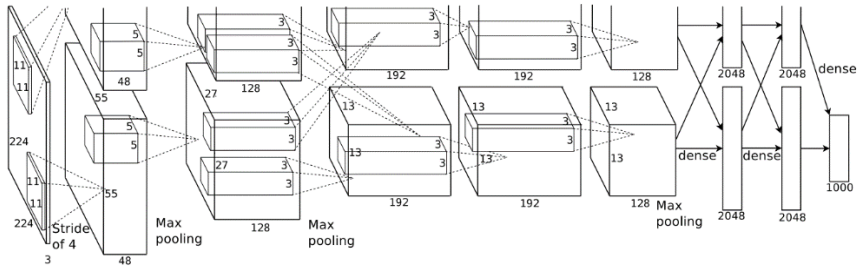
2.3. Mạng huấn luyện AlexNet

Mạng huấn luyện AlexNet là công trình đầu tiên phổ biến mạng CNN trong lĩnh vực Thị giác máy tính, cũng là một trong những mạng huấn luyện CNN nổi tiếng nhất nhờ thành tích ấn tượng mà nó đạt được trong cuộc thi nhận dạng ảnh quy mô lớn (ILSV) tổ chức vào năm 2012.

Nhiệm vụ chính của cuộc thi đặt ra mà các đội tham gia phải giải quyết là bài toán nhận dạng, với bộ dữ liệu huấn luyện lên đến 1,2 triệu ảnh được gán nhãn cho 1.000 hạng mục khác nhau. Nhóm SuperVision, gồm các thành viên Alex Krizhevsky, Ilya Sutskever và Geoff Hinton, cùng với mạng AlexNet của họ đã đạt được kết quả đáng kinh ngạc là chiến thắng áp đảo nhóm đứng thứ hai với độ chính xác chênh lệch đến hơn 10% (15,31% và 26,17%). Điều kinh ngạc là mạng huấn luyện này chỉ nhận dữ liệu đầu vào là các giá trị điểm ảnh thô và không hề áp dụng bất kỳ phương pháp trích chọn đặc trưng nào, trong khi mọi hệ thống nhận giác thị giác truyền thống đều phải gồm nhiều giai đoạn trích chọn đặc trưng hết sức tỉ mỉ, cẩn thận, thậm chí phải áp dụng nhiều mẹo để cải thiện chất lượng nhận dạng.

2.2.1. Kiến trúc mạng AlexNet

Nhóm của Alex Krizhevsky đã công bố một bài báo với tiêu đề “ImageNet Classification with Deep Convolutional Networks”, đưa ra mô tả cụ thể về kiến trúc của mạng AlexNet cũng như cách thức cài đặt và sử dụng các lớp trong mạng để huấn luyện mô hình với bộ dữ liệu ảnh khổng lồ của ImageNet. Mạng có cấu trúc tương đối đơn giản nếu so với các mạng CNN hiện đại gần đây, bao gồm 5 lớp Convolutional và 3 lớp Fully-connected, được huấn luyện song song trên 2 card đồ họa GPU.



2.4. Ứng dụng mạng AlexNet vào bài toán Nhận dạng, phân loại hoa quả

- 1) Cài đặt mạng AlexNet với một mô hình đã được huấn luyện trước với bộ ảnh của ImageNet.
- 2) Xây dựng bộ CSDL ảnh huấn luyện
- 3) Tinh chỉnh lại mô hình để giải quyết bài toán nhận dạng 40 loại hoa quả. Mặc dù kích thước CSDL ảnh không lớn, song độ chính xác của mô hình nhận dạng vẫn được đảm bảo nhờ khả năng trích chọn đặc trưng tự động của mạng AlexNet.

Chương 3. Tổng quan Hệ thống Fruit Recognition System

3.1. Tổng quan hệ thống

Hệ thống phần mềm Nhận dạng hoa quả – Fruit Recognition System – được thiết kế theo kiến trúc Client/Server năm tầng: Tầng Giao diện người dùng, Tầng Server quản lý cấu hình hệ thống, Tầng server thực hiện logic xử lý yêu cầu, Tầng huấn luyện mô hình nhận dạng và cuối cùng là Tầng quản lý CSDL.

Luồng hoạt động chính của hệ thống cụ thể như sau:

- **Client** (ứng dụng trên điện thoại thông minh):

- 1) Người dùng khởi động ứng dụng.
- 2) Người dùng thực hiện chụp ảnh hoa quả bằng camera, hoặc chọn ảnh đã chụp trước đó được lưu trong Thư viện ảnh.
- 3) Ảnh chụp được mã hóa, nén lại và gửi tới máy chủ.
- 4) Ứng dụng đợi nhận kết quả nhận dạng từ máy chủ gửi về và hiển thị cho người dùng.

- **Chương trình Server:**

- 1) Chương trình được khởi động và nạp các thư viện cần thiết.
- 2) Chương trình nạp mô hình nhận dạng đã được huấn luyện trước.
- 3) Giao thức gửi, nhận dữ liệu giữa ứng dụng phía client và chương trình server được cấu hình.
- 4) Một loại các luồng xử lý được khởi tạo, đặt trạng thái ban đầu là trạng thái rỗi.

- 5) Khi có ứng dụng client kết nối tới, chương trình kiểm tra trong danh sách các luồng xử lý và chọn một luồng đang ở trạng thái rồi để nhận và tính toán dữ liệu do client gửi tới.
- 6) Trong luồng xử lý:
 - Bắt đầu quá trình tính toán nhận dạng, cờ trạng thái là “bận”.
 - Thực hiện giải nén dữ liệu thành dữ liệu ảnh gốc.
 - Sử dụng mô hình đã nạp để nhận dạng loại hoa quả.
 - Trả kết quả nhận dạng về cho ứng dụng client.
 - Kết thúc quá trình tính toán.
- 7) Khi luồng xử lý đã hoàn thành quá trình tính toán nhận dạng, chương trình giải phóng luồng xử lý bằng cách cập nhật lại trạng thái hiện tại của luồng.

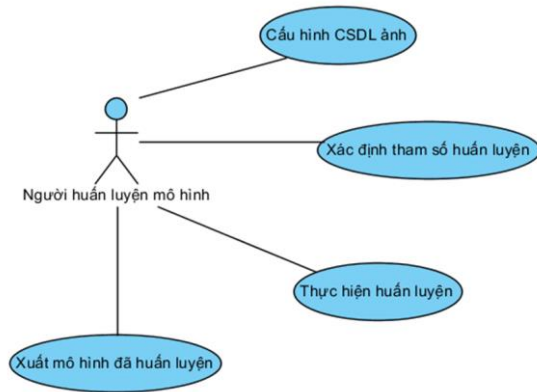
3.2. Mô đun quản lý cơ sở dữ liệu

Bộ CSDL ảnh phục vụ cho huấn luyện và tinh chỉnh các mô hình nhận dạng trong các thuật toán Học sâu nói riêng và Học máy nói chung là thành phần vô cùng quan trọng, quyết định chủ yếu đến độ chính xác mà mô hình đạt được. Do vậy, chúng cần được lưu trữ và quản lý một cách khoa học. Trong hệ thống lưu trữ, bộ CSDL ảnh huấn luyện được chia thành các thư mục riêng biệt: Thư mục ảnh gốc, Thư mục ảnh thực tế chưa duyệt và Thư mục ảnh thực tế đã duyệt.

3.3. Bộ huấn luyện mô hình

Nằm ở tầng thứ tư trong kiến trúc n tầng của hệ thống, bộ huấn luyện mô hình là thành phần có vai trò quan trọng hàng đầu, chịu toàn bộ trách nhiệm về các mô hình nhận dạng từ giai đoạn khởi tạo đến tinh chỉnh và hoàn thiện, cũng như quản lý và đánh giá độ chính xác

các phiên bản khác nhau của mô hình. Bộ huấn luyện được cài đặt và triển khai thành một mô đun hoàn toàn tách biệt với các thành phần còn lại của server, giúp cho việc nâng cấp hay thay thế có thể thực hiện độc lập mà không gây ảnh hưởng đến hoạt động thông thường của server.



3.3.1. Môi trường huấn luyện

Môi trường được nhóm đề tài sử dụng để huấn luyện mô hình nhận dạng hoa quả là Windows 10, ngôn ngữ Python phiên bản 2.7.12 với framework chuyên dùng cho Học sâu là Caffe. Caffe là một framework mã nguồn mở cho Học sâu, phát triển với Berkeley Vision and Learning Center (BVLC), được viết bởi ngôn ngữ C++, CUDA C++ cùng với các bộ gói wrapper cho các ngôn ngữ khác như Python hay Matlab. Điểm mạnh của framework này là cho phép người dùng tùy chọn huấn luyện thuật toán Học sâu trên CPU hay trên card đồ họa GPU, dễ dàng thực hiện quá trình huấn luyện trên bộ dữ liệu ảnh cá nhân chỉ với các câu lệnh đơn giản. Bên cạnh đó, Caffe cũng cho phép

người dùng tái sử dụng lại các mô hình đã được huấn luyện sẵn và được chia sẻ bởi cộng đồng nghiên cứu trên khắp thế giới.

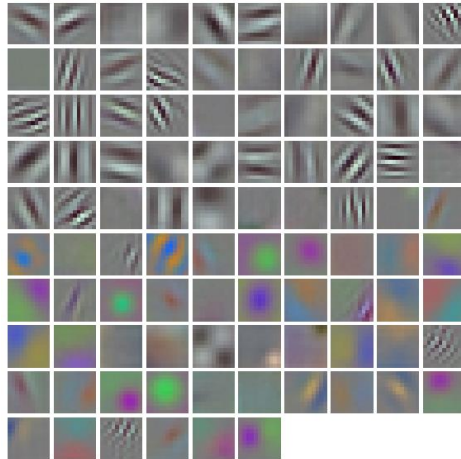
3.3.2. Cấu hình mạng huấn luyện AlexNet

Các mô hình huấn luyện và các phép tính chính của mạng AlexNet nói riêng và mạng CNN nói chung đều được thể hiện bằng cấu trúc văn bản thuần, nhằm tạo ra sự minh bạch rõ ràng khi định nghĩa các phép biến đổi ảnh hay các lớp trong mô hình và sự dễ dàng khi triển khai hoặc chuyển giao một mô hình huấn luyện.

Các thông số của từng lớp trong mạng huấn luyện được định nghĩa rõ ràng trong file cấu hình, để giúp framework thực hiện việc huấn luyện mô hình một cách chính xác. Tuy nhiên đối với trường hợp việc tinh chỉnh mô hình đã được huấn luyện trước, ta chỉ cần quan tâm tới các tham số chính như sau: Thông tin ảnh đầu vào, Thông số tỉ lệ học và Số lượng lớp nhận dạng đầu ra.

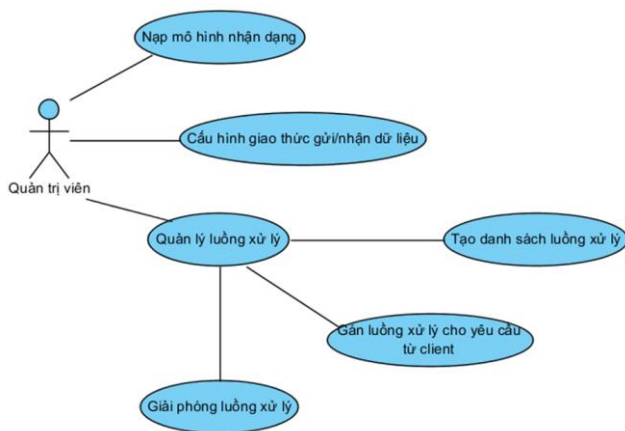
3.3.3. Một số hình ảnh về đặc trưng do mạng AlexNet tính toán

Các lớp đầu tiên trong mạng huấn luyện của CNN có thể được coi là bộ bóc tách các đặc trưng cơ bản, chung nhất cho mọi tác vụ nhận dạng ảnh như các đường thẳng, đường cong hay các đốm ảnh màu... Với Caffe ta hoàn toàn có thể hiển thị các đặc trưng này lên để hiểu rõ hơn các bước nâng dần mức trừu tượng qua từng lớp của mạng CNN. Ngoài ra độ nét và mịn của các đặc trưng cũng thể hiện cho mức độ huấn luyện của mạng, nếu mạng chưa được huấn luyện tốt, với kích thước CSDL ảnh lớn và thời gian huấn luyện đủ lâu, thì hình ảnh các đặc trưng sẽ bị nhiễu.



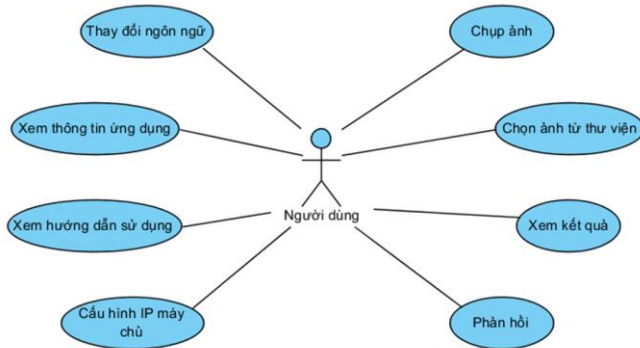
3.3.4. Các mô đun phía Server

Chương trình phía server được cấu thành bởi các nhiều mô đun, đảm nhiệm các vai trò nhiệm vụ khác nhau liên quan đến huấn luyện, quản lý mô hình nhận dạng, cấu hình giao thức giao tiếp giữa client và server hay xử lý logic đa luồng, đảm bảo khả năng tính toán cho nhiều yêu cầu cùng lúc...



3.4. Ứng dụng phía Client

Ứng dụng phía Client là ứng dụng trên điện thoại thông minh, là một thành phần trong hệ thống đảm nhiệm vai trò thu thập ảnh đầu vào để nhận dạng, gồm các chức năng chính sau đây: Chụp ảnh, Chọn ảnh từ thư viện và Xem kết quả nhận dạng do mô đun phía Server trả về.



Để kết quả nhận dạng được tốt, các ảnh đầu vào cần phải thỏa mãn một số ràng buộc chính, các ràng buộc này nhằm đảm bảo sự tương tự nhất định giữa ảnh đầu vào và bộ ảnh dữ liệu được sử dụng để huấn luyện, từ đó nâng cao tỉ lệ nhận dạng chính xác của mô hình.

- 1) Trong ảnh đầu vào chỉ có duy nhất một loại quả.
- 2) Hình ảnh của quả trong ảnh phải chiếm tỉ lệ nhất định trong ảnh, nếu hình ảnh của quả quá nhỏ sẽ dẫn đến khó khăn trong tính toán đặc trưng, gây nhầm lẫn giữa quả và nền và làm sai kết quả.
- 3) Hình ảnh của quả trong ảnh không bị che lấp quá nhiều bởi vật thể khác.
- 4) Ảnh chụp đầu vào không bị quá nhòe hoặc điều kiện ánh sáng quá kém.

Chương 4. Kết quả thử nghiệm và đánh giá

4.1. So sánh với phương pháp Machine Learning truyền thống

Thực nghiệm với phương pháp Học máy truyền thống:

- **Bước 1:** Xây dựng CSDL ảnh hoa quả cho 20 loại quả.
- **Bước 2:** Tiền xử lý ảnh trong CSDL (lọc nền) và gán nhãn. Đặc trưng của bộ CSDL ảnh này là các ảnh đều có chất lượng cao, có cùng kích thước và tỉ lệ ảnh, với nền đã bị loại bỏ hoàn toàn.



- **Bước 3:** Chọn lọc đặc trưng, cụ thể:
 - Về màu sắc: Sử dụng 16 đặc trưng về số lượng các điểm ảnh có giá trị màu tính theo hệ màu HSI (Hue-Saturation-Intensity).
 - Về hình dạng: Sử dụng 4 đặc trưng về hình dạng của hoa quả trong ảnh là chu vi, diện tích, độ dài lớn nhất, độ rộng lớn nhất của hoa quả trong ảnh.
 - Về kết cấu: Sử dụng 8 đặc trưng về kết cấu, là 8 tham số trong bộ ma trận GLCM (Grey Level Co-occurrence Matrix).
- **Bước 4:** Huấn luyện mô hình nhận dạng hoa quả.
- **Bước 5:** Thống kê độ chính xác với tỉ lệ bộ training/test là 75/25. Kết quả đạt được không cao, chỉ đạt ~74.5%

trên bộ dữ liệu test 2.600 ảnh, và khi thử nghiệm thực tế cũng gặp phải sai số lớn (do ảnh chụp thực tế có chất lượng không cao và sự khác biệt lớn so với bộ CSDL ảnh để huấn luyện).

Thực nghiệm với phương pháp Học sâu:

- **Bước 1:** Xây dựng CSDL ảnh hoa quả cho 20 loại quả.
- **Bước 2:** Tiền xử lý ảnh trong CSDL (lọc nền) và gán nhãn. Hai bước đầu tiên này chỉ cần thực hiện một lần khi xây dựng bộ CSDL ảnh huấn luyện cho phương pháp Học máy truyền thống.
- **Bước 3:** Thực hiện các bước tính toán cần thiết để ứng dụng mô hình AlexNet.
- **Bước 4:** Huấn luyện mô hình nhận dạng hoa quả.
- **Bước 5:** Thống kê độ chính xác với tỉ lệ bộ training/test là 75/25. Độ chính xác đạt được là rất cao, ~98.8%, vượt trội so với độ chính xác của mô hình huấn luyện sử dụng phương pháp Học máy truyền thống.

Đánh giá kết quả: Với các bài toán nhận dạng và phân loại đối tượng nói chung, trong đó rất khó có thể chọn được các đặc trưng hiệu quả, thì Học sâu là phương pháp có ưu thế vượt trội so với các phương pháp Học máy truyền thống. Học sâu giúp đơn giản hóa quá trình huấn luyện mô hình nhận dạng khi không yêu cầu sự tham gia của người huấn luyện trong quá trình trích chọn đặc trưng, đồng thời cho phép tái sử dụng các mô hình đã huấn luyện trước để giảm thời gian cài đặt giải pháp cho các bài toán nhận dạng mới.

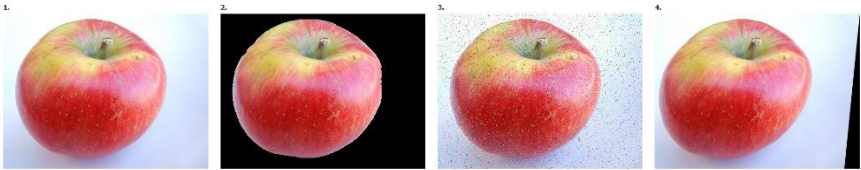
4.2. So sánh kết quả với bộ CSDL được sinh tự động

Thực nghiệm với bộ CSDL ảnh gốc:

- Số lượng ảnh gốc cho mỗi loại quả: 500-1000 ảnh
- Tổng số ảnh dùng để huấn luyện: 21.000 ảnh
- Tổng số ảnh dùng để test: 7.000 ảnh
- Thời gian huấn luyện cho 10.000 vòng lặp: 5 tiếng
- Độ chính xác: 65,49%

Thực nghiệm với bộ CSDL ảnh được sinh tự động từ ảnh gốc:

Từ mỗi ảnh gốc, sau khi sử dụng các thuật toán xử lý ảnh như chiếu nghiêng (skew), thêm nhiễu và ghép nền khác ta sẽ thu được 9 ảnh mới để tăng cường cho bộ CSDL ảnh huấn luyện.



- Số lượng ảnh gốc cho mỗi loại quả: 500-1000 ảnh
- Số lượng ảnh sinh thêm từ một ảnh gốc: 9 ảnh
- Tổng số ảnh dùng để huấn luyện: 210.000 ảnh
- Tổng số ảnh dùng để test: 70.000 ảnh
- Thời gian huấn luyện cho 10.000 vòng lặp: 50 tiếng
- Độ chính xác: 98,67%

Đánh giá kết quả: Sự cải thiện rõ rệt trong độ chính xác của mô hình nhận dạng sau khi tăng cường CSDL ảnh huấn luyện đã cho thấy hiệu quả thực tế của các phép sinh ảnh tự động sử dụng các phương pháp xử lý ảnh cơ bản. Chất lượng nhận dạng của ứng dụng trong thực tế cũng được tăng lên do các ảnh được sinh tự động giúp mô phỏng quá trình chụp ảnh trong đời thực,

nhu các góc chụp khác nhau, các nhiễu sinh ra do môi trường, chất lượng máy ảnh... cũng như sự đa dạng của nền mà người dùng sử dụng để chụp ảnh. Việc tăng cường CSDL ảnh cũng là một giải pháp cho trường hợp khó thu thập ảnh để huấn luyện mô hình, tuy nhiên cũng cần phải chú ý đến mặt trái của việc lạm dụng phương pháp tăng cường ảnh này, đó là nguy cơ gây ra trạng thái “overfit” dữ liệu (mô hình nhận dạng quá khớp với dữ liệu huấn luyện mà bị sai lệch với dữ liệu thực tế).

4.3. Thử nghiệm ứng dụng trong thực tế

Ứng dụng Nhận dạng hoa quả đã được thử nghiệm thực tế với nhiều mẫu hoa quả khác nhau, được chia thành hai nhóm chính: Nhóm đã được huấn luyện nhận dạng và nhóm chưa được huấn luyện.

- Nhóm hoa quả nằm trong danh sách 40 quả được chọn để xây dựng bộ dữ liệu nhận dạng: Kết quả nhận dạng đạt độ chính xác khá cao, đặc biệt là với những loại quả có nét đặc trưng về màu sắc hoặc hình dạng như chuối, thanh long, chôm chôm...



```
C:\Windows\System32\cmd.exe - python FruitServer3.py
Nhập đường link ảnh
C:\Users\Dell\Pictures\Master\banana.png
C:\Miniconda2\lib\site-packages\skimage\transform\_warps.py:84:
.15.
warn("The default mode, 'constant', will be changed to 'refl
predicted class is: 3
----- 0.987864 ['3' 'banana' 'Chuoi']
----- 0.0107678 ['20' 'mango' 'Xoai']
----- 0.00063867 ['22' 'orange' 'Cam']
----- 0.000308139 ['23' 'papaya' 'Du Du']
----- 9.61382e-05 ['36' 'starfruit' 'Khe']
Nhập đường link ảnh
```

Đối với những loại quả có nhiều nét tương đồng lẫn nhau, kết quả nhận dạng của ứng dụng còn đôi lúc bị nhầm lẫn.



```
C:\Windows\System32\cmd.exe - python FruitServer3.py
Nhap duong link anh
C:\Users\Dell\Pictures\Master\peach2.png
predicted class is: 29
----- 0.416721 ['29' 'plum' 'So Ri']
----- 0.342792 ['25' 'peach' 'Dao']
----- 0.219008 ['20' 'mango' 'Xoai']
----- 0.00672279 ['27' 'persimmon' 'Hong']
----- 0.00473678 ['0' 'apple' 'Tao']
Nhap duong link anh
```

- Nhóm hoa quả nằm ngoài danh sách 40 quả: Hệ thống sẽ tính toán và trả về kết quả nhận dạng là một trong 40 loại hoa quả có tỉ lệ giống nhất với loại quả cần nhận dạng.



```
C:\Windows\System32\cmd.exe - python FruitServer3.py
----- 0.00063867 ['22' 'orange' 'Cam']
----- 0.000308139 ['23' 'papaya' 'Du Du']
----- 9.61382e-05 ['36' 'starfruit' 'Khe']
Nhap duong link anh
C:\Users\Dell\Pictures\Master\bonbon.png
predicted class is: 18
----- 0.99341 ['18' 'longan' 'Nhan']
----- 0.00494929 ['33' 'sapota' 'Hong Xiem']
----- 0.000585203 ['6' 'coconut' 'Dua']
----- 0.000389234 ['14' 'honeydew melon' 'Dua Gang']
----- 0.000253315 ['19' 'lychee' 'Vai']
Nhap duong link anh
```

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 ta đã đề cập tới trong phần mở đầu của luận văn, 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.

Chương 5. Kết luận

Luận văn đã nghiên cứu, tìm hiểu bài toán tự động nhận dạng, phân loại hoa quả trong ảnh màu, và thực hiện phát triển, cài đặt phương án giải quyết cho bài toán dựa trên sự tìm kiếm, 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 ban đầu như sau:

- Hoàn thiện xây dựng bộ cơ sở dữ liệu ảnh phục vụ huấn luyện nhận dạng cho 40 loại hoa quả phổ biến ở nước ta, với số lượng ảnh gốc trung bình cho mỗi loại quả là từ 500-1000 ảnh.

- Thống kê các đặc trưng thường được sử dụng để huấn luyện bộ nhận dạng hoa quả trong các phương pháp Học máy truyền thống, bao gồm các đặc trưng về màu sắc, hình dạng và kết cấu. Từ đó làm cơ sở xây dựng một mạng nơ-ron nhân tạo truyền thống và so sánh kết quả với một mạng nơ-ron tích chập thuộc nhóm phương pháp Học sâu.

- Cài đặt và tinh chỉnh một mạng nơ-ron tích chập đã được huấn luyện trước, ứng dụng vào bài toán nhận dạng hoa quả. Đồng thời xây dựng hệ thống tự động nhận dạng hoa quả Fruit Recognition System với ứng dụng client trên smartphone.

Thực nghiệm với bộ dữ liệu test và trong thực tế đã cho kết quả khá tốt, nguyên nhân chính là do phạm vi số lượng hoa quả để nhận dạng đã được hạn chế chỉ còn 40 loại – một con số rất khiêm tốn so với số lượng hoa quả ở Việt Nam nói riêng và cả thế giới nói chung. Hệ thống tự động nhận dạng hoa quả còn cần rất nhiều cải thiện, đặc biệt là về khả năng mở rộng phạm vi loại hoa quả cũng như kích thước,

chất lượng của bộ CSDL ảnh huấn luyện. Trong tương lai, để có thể cải thiện độ chính xác của mô hình nhận dạng, nhóm đề tài đề xuất cài đặt thử nghiệm và đánh giá các loại mô hình mạng Học sâu đã được huấn luyện trước, đặc biệt là các mạng đã đạt được kết quả cao trong cuộc thi Nhận dạng ảnh quy mô lớn do ImageNet tổ chức thường niên như: ZF Net (2013), VGG Net (2014), GoogleNet và Microsoft ResNet (2015)...