

ỦY BAN NHÂN DÂN TP. HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**---------------o0o---------------**

**ĐỀ TÀI : NHẬN DIỆN CÁC LOẠI TRÁI CÂY BẰNG HÌNH ẢNH DÙNG PHƯƠNG PHÁP DEEP LEARNING**

**MÔN HỌC : PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC**

**GIẢNG VIÊN HƯỚNG DẪN :** TS. ĐỖ NHƯ TÀI

**NHÓM SINH VIÊN THỰC HIỆN :**

1. TRẦN HOÀNG KHANH\_3121410257
2. DƯƠNG DUY KHANG\_3121410250
3. NGUYỄN NGỌC HẢI\_3121410175

**Thành phố Hồ Chí Minh,** **<Tháng>/<Năm>**

**MỤC LỤC**

[**TÓM TẮT 5**](#_xxlcdgsavcmw)

[**Chương 1 : Tổng quan vấn đề 8**](#_qq6sx3r87gsw)

[1.1. Lý do chọn đề tài 8](#_1io8tgfyu4v5)

[1.2. Vấn đề nghiên cứu 8](#_x2jsgb7fmfwh)

[1.3. Mục đích nghiên cứu 8](#_fqmsz6c4t0pt)

[1.4. Câu hỏi nghiên cứu 9](#_syxk43ssbcsj)

[1.5. Phạm vi nghiên cứu 10](#_hlz10ofrb4go)

[**Chương 2 : Lược khảo tài liệu (Literature Review) 11**](#_nmnto3mcvsi)

[2.1. Tổng hợp các tài liệu, nghiên cứu trước liên quan. 11](#_ia65wrvd00fd)

[2.2. Cơ sở lý thuyết cho nghiên cứu 11](#_vddwhv54xj9)

[2.2.1. Thị giác máy tính (Computer Vision) 11](#_b2in7moi013h)

[2.2.2. Học sâu (Deep Learning) và Mạng nơ-ron tích chập (CNN) 11](#_b4wtd1o4ax7y)

[2.2.3. Tập dữ liệu Fruits-360. 12](#_be0v6jx1y0kg)

[2.3. Nêu điểm mạnh, điểm yếu của các nghiên cứu trước và cách nghiên cứu của bạn kế thừa hoặc phát triển. 13](#_x5k2qstvzhy6)

[**Chương 3: Phương pháp nghiên cứu (Methodology) 15**](#_iefecohxtq45)

[3.1. Cấu trúc của mạng nơ-ron được sử dụng trong các thí nghiệm 15](#_t15gmz6q3f0s)

[3.2 Đối tượng và mẫu nghiên cứu 17](#_etc2d5vlawlt)

[3.3. Cách thu thập dữ liệu 21](#_oba6r9s4tu52)

[3.4 Thư viện TensorFlow. 22](#_ciooza6u9uhi)

[**Chương 4. Thực nghiệm và Thảo luận 25**](#_us50mtsi5rhy)

[4.1. Thí nghiệm số. 25](#_rm6nzb12ul2h)

[4.2 Đào tạo mạng lưới thần kinh 25](#_yyd66g8f814r)

[4.3. Đào tạo các cấu hình mạng khác nhau. 27](#_w1sc5mm1pp0l)

[4.4 Sai số thực nghiệm. 29](#_lfnvbsuvz65l)

[**Chương 5. Kết luận và Hướng phát triển 31**](#_dbauwvdlvp75)

[5.1. Tóm tắt kết quả đạt được và hướng nghiên cứu tương lai. 31](#_670mn430qpu3)

[5.2. Câu trả lời cho câu hỏi nghiên cứu 32](#_v4bb10ipf0b6)

[**TÀI LIỆU THAM KHẢO 33**](#_3s0m2l20j5rd)

[**PHỤ LỤC 37**](#_8xdc20l9ccrl)

**DANH MỤC BẢNG BIỂU, HÌNH ẢNH**

Hình ảnh minh họa, bảng biểu…được ghi theo ký tự của từng chương.

Ví dụ:

Hình 1 của chương 1 : Hình 1.1. tên hình

Hình 1 của chương 2 : Hình 2.1. tên hình

Bảng 1 của chương 3 . Bảng 3.1. tên bảng

**DANH MỤC CÁC CHỮ VIẾT TẮT**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# TÓM TẮT

**Lý do nghiên cứu**

Lý do nghiên cứu "Nhận diện và phân loại trái cây thông qua hình ảnh bằng phương pháp học sâu" xuất phát từ nhu cầu phân loại tự động, thay thế phương pháp nhận diện thủ công vốn tốn thời gian và dễ sai sót. Machine Learning giúp xử lý hình ảnh nhanh chóng, cải thiện độ chính xác, giúp xác định loại trái cây một cách hiệu quả. Việc nghiên cứu này nhằm tối ưu hóa mô hình nhận diện, góp phần nâng cao tốc độ xử lý và áp dụng công nghệ AI vào lĩnh vực phân loại thực phẩm và tự động hóa.

**Mục tiêu nghiên cứu**

Mục tiêu nghiên cứu của đề tài "Nhận diện và phân loại trái cây thông qua hình ảnh bằng phương pháp học sâu" nhằm xây dựng một hệ thống tự động nhận diện và phân loại trái cây thông qua hình ảnh, góp phần tối ưu hóa quy trình kiểm tra chất lượng, phân loại sản phẩm và tự động hóa trong nông nghiệp. Cụ thể:

* Phát triển mô hình phương pháp học sâu có khả năng phân loại các loại trái cây dựa trên hình ảnh với độ chính xác cao.
* Nghiên cứu và tối ưu hóa thuật toán xử lý ảnh, đặc biệt là mạng nơ-ron tích chập (CNN) để nâng cao hiệu suất nhận diện.
* Đánh giá hiệu suất của mô hình, đảm bảo khả năng nhận diện chính xác các loại trái cây khác nhau trong tập dữ liệu.
* So sánh độ hiệu quả của các phương pháp nhận diện, để xác định thuật toán phù hợp nhất cho bài toán nhận diện trái cây.

Nghiên cứu này giúp nâng cao độ chính xác và tốc độ nhận diện, góp phần tối ưu hóa các quy trình kiểm tra sản phẩm, đồng thời mở ra hướng phát triển mới cho AI trong lĩnh vực nông nghiệp

**Phương pháp nghiên cứu**

Ứng dụng **phương pháp học sâu** để xây dựng mô hình phân loại trái cây dựa trên hình ảnh. Điều này giúp quá trình sắp xếp và phân loại trở nên tự động hóa gần như hoàn toàn.

Input: Ảnh của một loại trái cây trên nền đơn giản (một màu).

Output: Tên của loại trái cây đó.

**Kết quả chính ( nếu có )**

Nghiên cứu về ứng dụng **phương pháp học sâu** để nhận diện trái cây nhằm xây dựng một hệ thống **tự động phân loại trái cây dựa trên hình ảnh**, giúp thay thế phương pháp nhận diện thủ công vốn tốn thời gian và dễ sai sót. Sử dụng mạng nơ-ron tích chập (CNN) và các thuật toán xử lý ảnh, mô hình được phát triển nhằm tối ưu hóa **độ chính xác khi nhận diện**, đảm bảo khả năng phân loại đúng nhiều loại trái cây khác nhau. Kết quả nghiên cứu đã chỉ ra rằng việc áp dụng Machine Learning giúp **tăng tốc độ xử lý, giảm sai sót**, và **mở rộng khả năng nhận diện trên tập dữ liệu lớn**. Đồng thời, việc đánh giá hiệu suất giữa các thuật toán như CNN, SVM, và Random Forest đã xác định được phương pháp phù hợp nhất để giải quyết bài toán nhận diện trái cây. Nghiên cứu này **không mở rộng sang các lĩnh vực khác**, mà tập trung hoàn toàn vào việc nhận diện trái cây nhằm đảm bảo hệ thống hoạt động hiệu quả và chính xác

**Kết luận chính**

Nghiên cứu về ứng dụng **phương pháp học sâu** để nhận diện trái cây đã đạt được mục tiêu xây dựng một hệ thống tự động phân loại trái cây dựa trên hình ảnh, giúp thay thế phương pháp nhận diện thủ công vốn tốn thời gian và dễ sai sót. **Phương pháp học sâu**, đặc biệt là mạng nơ-ron tích chập (CNN), đã được chứng minh là phương pháp hiệu quả, giúp cải thiện độ chính xác, tăng tốc độ xử lý và đảm bảo khả năng phân loại đúng nhiều loại trái cây khác nhau. Quá trình thử nghiệm đã cho thấy sự khác biệt về hiệu suất giữa các thuật toán như CNN, SVM và Random Forest, từ đó xác định mô hình phù hợp nhất cho bài toán nhận diện trái cây.

Kết quả nghiên cứu đã khẳng định rằng việc áp dụng **phương pháp học sâu** trong nhận diện trái cây giúp tối ưu hóa quy trình phân loại, tạo tiền đề cho việc ứng dụng AI vào tự động hóa nông nghiệp. Tuy nhiên, đề tài chỉ tập trung vào nhận diện trái cây, không mở rộng sang các lĩnh vực khác, nhằm đảm bảo độ chính xác và hiệu quả của hệ thống.

# Chương 1 : Tổng quan vấn đề

## 1.1. Lý do chọn đề tài

Lý do chọn đề tài "Nhận diện và phân loại trái cây thông qua hình ảnh bằng phương pháp học sâu" xuất phát từ nhu cầu tự động hóa quá trình phân loại trái cây, thay thế phương pháp nhận diện thủ công vốn dễ sai sót và mất nhiều thời gian. Việc áp dụng Machine Learning giúp mô hình nhận diện chính xác hơn, xử lý nhanh hơn, đồng thời đảm bảo khả năng phân loại đúng nhiều loại trái cây khác nhau. Nghiên cứu này tập trung hoàn toàn vào việc phát triển thuật toán nhận diện, không mở rộng sang các lĩnh vực khác, nhằm đảm bảo hệ thống hoạt động hiệu quả và chính xác nhất. Kết quả đạt được sẽ giúp tối ưu hóa công nghệ nhận diện trái cây, hướng đến việc ứng dụng thực tế trong tự động hóa nông nghiệp.

## 1.2. Vấn đề nghiên cứu

Trong thực tế, việc nhận diện trái cây bằng phương pháp thủ công thường **tốn thời gian và dễ sai sót**, đặc biệt trong các hệ thống tự động hóa sản xuất và kiểm tra chất lượng thực phẩm. Vì vậy, cần xây dựng một mô hình **tự động phân loại trái cây** với độ chính xác cao, giúp tối ưu hóa quy trình phân loại sản phẩm.

* Input : Ảnh của một loại trái cây trên nền đơn giản (một màu). ​
* Output :Tên của loại trái cây đó.

## 1.3. Mục đích nghiên cứu

Nghiên cứu **"Nhận diện và phân loại trái cây thông qua hình ảnh bằng phương pháp học sâu"** dựa trên tập dữ liệu **Fruits-360**, nhằm phát triển một mô hình **Deep Learning** có khả năng **tự động nhận diện và phân loại trái cây** với độ chính xác cao. Việc phân loại trái cây bằng phương pháp thủ công thường mất nhiều thời gian và dễ xảy ra sai sót, đặc biệt trong các hệ thống **kiểm tra chất lượng và tự động hóa sản xuất**. Do đó, ứng dụng **mạng nơ-ron tích chập (CNN)** và các kỹ thuật học sâu giúp tối ưu hóa quá trình này, nâng cao hiệu suất và đảm bảo phân loại chính xác từng loại trái cây.

Cụ thể, nghiên cứu này hướng đến:

**Xây dựng mô hình CNN** có khả năng nhận diện từng loại trái cây dựa trên hình ảnh trong tập dữ liệu Fruits-360.

**Tối ưu hóa thuật toán xử lý ảnh**, giúp mô hình hoạt động ổn định và chính xác dù điều kiện ánh sáng hoặc góc chụp khác nhau.

**Đánh giá độ chính xác của mô hình**, đảm bảo hiệu suất nhận diện trên tập kiểm tra.

**Ứng dụng vào thực tế**, hướng đến tích hợp hệ thống nhận diện vào tự động hóa nông nghiệp và kiểm tra chất lượng trái cây.

Nghiên cứu này **chỉ tập trung vào nhận diện và phân loại trái cây**, không mở rộng sang các lĩnh vực khác, nhằm đảm bảo **độ chính xác cao nhất** trong bài toán đặt ra.

## 1.4. Câu hỏi nghiên cứu

**Câu hỏi nghiên cứu tập trung vào nhận diện trái cây:**

1. **Mô hình học sâu nào hiệu quả nhất** để nhận diện trái cây dựa trên tập dữ liệu Fruits-360?
2. **Kỹ thuật tiền xử lý hình ảnh nào** giúp tối ưu hóa mô hình nhận diện trái cây?
3. **Độ chính xác của CNN khi nhận diện từng loại trái cây** đạt mức nào, và có cần cải tiến thêm không?
4. **Yếu tố nào ảnh hưởng đến khả năng nhận diện**, ví dụ như kích thước, hình dạng, ánh sáng và nền ảnh?
5. **Làm thế nào để đảm bảo mô hình có thể nhận diện chính xác các loại trái cây khác nhau trong điều kiện thực tế?**

## 1.5. Phạm vi nghiên cứu

Nghiên cứu này tập trung vào **ứng dụng Machine Learning để nhận diện trái cây**, sử dụng tập dữ liệu **Fruits-360** gồm hơn **130.000 hình ảnh của 194 loại trái cây và rau củ**. Phạm vi nghiên cứu được xác định như sau:

**Đối tượng nghiên cứu:** Các thuật toán Machine Learning, đặc biệt là **mạng nơ-ron tích chập (CNN)**, được sử dụng để **phân loại trái cây dựa trên hình ảnh** trong tập dữ liệu Fruits-360.

**Dữ liệu nghiên cứu:** Chỉ sử dụng hình ảnh của trái cây từ **Fruits-360**, đảm bảo tính thống nhất và chất lượng dữ liệu phù hợp cho quá trình huấn luyện mô hình.

**Phương pháp nghiên cứu:** Tiền xử lý hình ảnh, huấn luyện mô hình CNN, và đánh giá hiệu suất trên tập kiểm tra từ cùng tập dữ liệu Fruits-360.

**Giới hạn nghiên cứu:**

* Không mở rộng nghiên cứu sang **nhận diện thực phẩm khác ngoài trái cây**.
* Không nghiên cứu **tích hợp mô hình vào hệ thống kiểm tra chất lượng thực tế**, chỉ tập trung vào **phân loại hình ảnh**.
* Không sử dụng **dữ liệu bên ngoài** ngoài Fruits-360 để đảm bảo tính nhất quán trong nghiên cứu.

Với phạm vi này, nghiên cứu sẽ tập trung hoàn toàn vào **xây dựng, huấn luyện và tối ưu hóa mô hình nhận diện trái cây**, đảm bảo hệ thống hoạt động hiệu quả trên tập dữ liệu đã được chọn.

# Chương 2 : Lược khảo tài liệu (Literature Review)

## 2.1. Tổng hợp các tài liệu, nghiên cứu trước liên quan.

* **HGCLIP: Exploring Vision-Language Models with Graph Representations for Hierarchical Understanding**, 23 Nov 2023,[Peng Xia](https://paperswithcode.com/author/peng-xia), [Xingtong Yu](https://paperswithcode.com/author/xingtong-yu), [Ming Hu](https://paperswithcode.com/author/ming-hu), [Lie Ju](https://paperswithcode.com/author/lie-ju), [Zhiyong Wang](https://paperswithcode.com/author/zhiyong-wang), [Peibo Duan](https://paperswithcode.com/author/peibo-duan), [ZongYuan Ge](https://paperswithcode.com/author/zongyuan-ge-1)
* **A Novel lightweight Convolutional Neural Network, ExquisiteNetV2**, 19 May 2021, Shi-Yao Zhou, Chung-Yen Su
* **Fruit and Vegetable Identification Using Machine Learning for Retail Applications,** 23 Oct 2018, Frida Femling, Adam Olsson, Fernando Alonso-Fernandez

## 2.2. **Cơ sở lý thuyết cho nghiên cứu**

Nghiên cứu **"Nhận diện và phân loại trái cây thông qua hình ảnh bằng phương pháp học sâu"** dựa trên tập dữ liệu **Fruits-360**, sử dụng nền tảng lý thuyết từ các lĩnh vực chính sau:

### **2.2.1. Thị giác máy tính (Computer Vision)**

Thị giác máy tính giúp máy có khả năng hiểu và phân tích hình ảnh, đặc biệt trong bài toán nhận diện trái cây. Các công nghệ quan trọng liên quan:

**Xử lý ảnh số**: Kỹ thuật tiền xử lý giúp chuẩn hóa hình ảnh, cân bằng ánh sáng, tăng cường dữ liệu để cải thiện hiệu suất mô hình.

**Trích xuất đặc trưng**: Xác định các yếu tố đặc trưng như màu sắc, hình dạng, kết cấu để mô hình dễ dàng phân loại trái cây.

### **2.2.2. Học sâu (Deep Learning) và Mạng nơ-ron tích chập (CNN)**

Trong lĩnh vực nhận dạng và phân loại hình ảnh, kết quả thành công nhất thu được bằng cách sử dụng mạng nơ-ron nhân tạo [6, 30]. Các mạng này tạo thành cơ sở cho hầu hết các mô hình học sâu.

Học sâu là một lớp thuật toán học máy sử dụng nhiều lớp chứa các đơn vị xử lý phi tuyến tính [26]. Mỗi cấp độ học cách biến đổi dữ liệu đầu vào của nó thành một biểu diễn trừu tượng và tổng hợp hơn một chút [6]. Mạng nơ-ron sâu đã cố gắng vượt trội hơn các thuật toán học máy khác. Chúng cũng đạt được khả năng nhận dạng mẫu siêu phàm đầu tiên trong một số lĩnh vực nhất định [5]. Điều này càng được củng cố bởi thực tế rằng học sâu được coi là một bước quan trọng để đạt được AI mạnh. Thứ hai, mạng nơ-ron sâu - cụ thể là mạng nơ-ron tích chập - đã được chứng minh là đạt được kết quả tuyệt vời trong lĩnh vực nhận dạng hình ảnh.

Mạng nơ-ron tích chập (CNN) là một phần của các mô hình học sâu. Một mạng như vậy có thể bao gồm các lớp tích chập, lớp gộp, các lớp ReLU, lớp kết nối đầy đủ và lớp mất mát [34]. Trong một kiến ​​trúc CNN điển hình, mỗi lớp tích chập được theo sau bởi một lớp Đơn vị tuyến tính chỉnh lưu (ReLU), sau đó là một lớp gộp, sau đó là một hoặc nhiều lớp tích chập và cuối cùng là một hoặc nhiều lớp kết nối đầy đủ. Một đặc điểm phân biệt CNN với mạng nơ-ron thông thường là tính đến cấu trúc

của hình ảnh trong khi xử lý chúng. Lưu ý rằng mạng nơ-ron thông thường chuyển đổi đầu vào thành một mảng một chiều khiến bộ phân loại đã được đào tạo ít nhạy cảm hơn với các thay đổi về vị trí

### 2.2.3. Tập dữ liệu Fruits-360.

Bộ dữ liệu kết quả có 61934 hình ảnh trái cây trải rộng trên 90 nhãn. Bộ dữ liệu có sẵn trên GitHub [18] và Kaggle [19]. Các nhãn và số lượng hình ảnh để đào tạo được đưa ra trong Bảng 1.

## 2.3. Nêu điểm mạnh, điểm yếu của các nghiên cứu trước và cách nghiên cứu của bạn kế thừa hoặc phát triển.

**Điểm mạnh:**

**Tập dữ liệu phong phú**: Fruits-360 cung cấp hơn **130.000 hình ảnh của 194 loại trái cây và rau củ**, giúp huấn luyện mô hình với dữ liệu đa dạng.

**Ứng dụng học sâu (Deep Learning)**: Nhiều nghiên cứu đã sử dụng **mạng nơ-ron tích chập (CNN)** để nhận diện trái cây với độ chính xác cao.

**Tiền xử lý dữ liệu tốt**: Các nghiên cứu trước đã áp dụng kỹ thuật **chuẩn hóa ảnh, tăng cường dữ liệu**, giúp cải thiện hiệu suất mô hình.

**Điểm yếu:**

**Giới hạn về kích thước ảnh**: Fruits-360 sử dụng ảnh **100x100 pixel**, có thể làm mất một số chi tiết quan trọng khi nhận diện.

**Thiếu dữ liệu thực tế**: Hầu hết ảnh trong tập dữ liệu là ảnh **được chụp trong điều kiện lý tưởng**, không phản ánh đầy đủ môi trường thực tế như ánh sáng yếu hoặc nền phức tạp.

**Chưa tối ưu hóa mô hình cho điều kiện thực tế**: Một số nghiên cứu chỉ tập trung vào **độ chính xác trên tập kiểm tra**, chưa đánh giá khả năng hoạt động của mô hình trong môi trường thực tế.

**Cách nghiên cứu này kế thừa và phát triển**

Dựa trên những điểm mạnh và điểm yếu của các nghiên cứu trước, nghiên cứu **"Nhận diện và phân loại trái cây thông qua hình ảnh bằng phương pháp học sâu"** sẽ kế thừa và phát triển theo hướng:

**Tận dụng CNN nhưng cải tiến kiến trúc**: Sử dụng **mô hình CNN nâng cao**, có thể kết hợp với **ResNet hoặc EfficientNet** để cải thiện độ chính xác.

**Tăng cường dữ liệu thực tế**: Áp dụng **kỹ thuật tăng cường dữ liệu**, như xoay ảnh, thay đổi độ sáng, làm mờ nền để mô hình hoạt động tốt hơn trong điều kiện thực tế.

**Đánh giá mô hình trên dữ liệu thực tế**: Thay vì chỉ kiểm tra trên tập Fruits-360, nghiên cứu sẽ thử nghiệm mô hình với **ảnh chụp thực tế**, giúp đánh giá khả năng nhận diện trong môi trường đa dạng.

Nghiên cứu này không chỉ kế thừa những thành tựu trước đó mà còn **khắc phục những hạn chế**, giúp mô hình nhận diện trái cây hoạt động hiệu quả hơn trong thực tế.

# Chương 3: Phương pháp nghiên cứu (Methodology)

## 3.1. Cấu trúc của mạng nơ-ron được sử dụng trong các thí nghiệm

Đối với dự án này, chúng tôi đã sử dụng mạng nơ-ron tích chập. Như đã mô tả trước đó, loại mạng này sử dụng các lớp tích chập, các lớp gộp, các lớp ReLU, các lớp được kết nối đầy đủ và các lớp mất mát. Trong một kiến ​​trúc CNN điển hình, mỗi lớp tích chập được theo sau bởi một lớp Đơn vị tuyến tính chỉnh lưu (ReLU), sau đó là một lớp Ghép, sau đó là một hoặc nhiều lớp tích chập và cuối cùng là một hoặc nhiều lớp được kết nối đầy đủ.

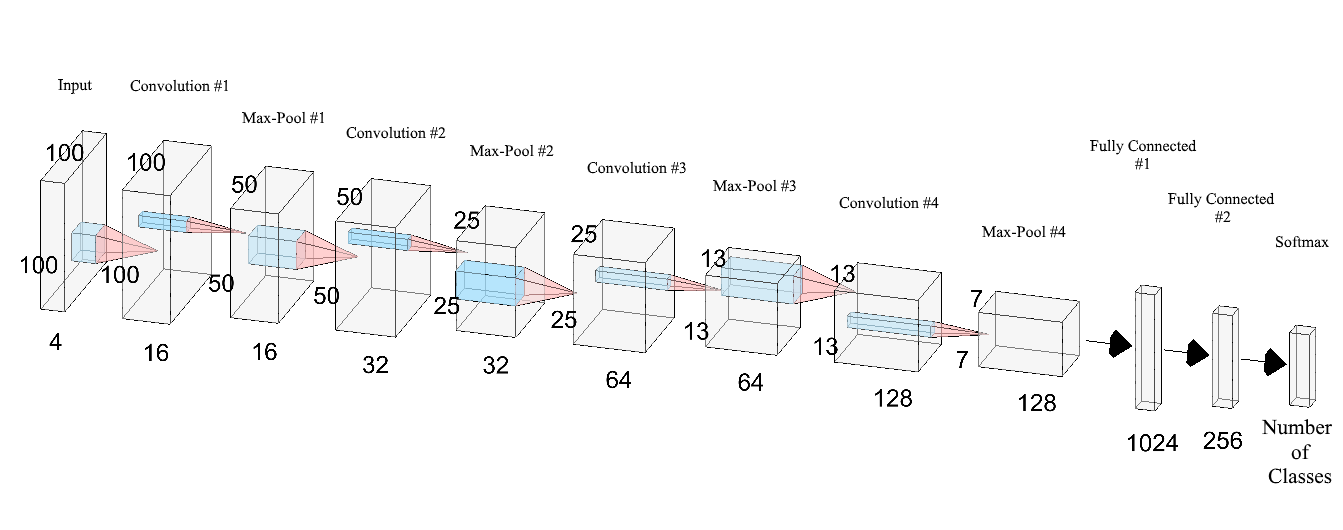
Đầu vào mà chúng tôi sử dụng bao gồm các hình ảnh RGB chuẩn có kích thước 100 x 100 pixel. Mạng nơ-ron mà chúng tôi sử dụng trong dự án này có cấu trúc được đưa ra trong Bảng 2.

Bảng 2: Cấu trúc của mạng nơ-ron được sử dụng trong bài báo này.

| Kiểu lớp | Kích thước | Đầu ra |
| --- | --- | --- |
| Convolutional | 5 x 5 x 4 | 16 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Convolutional | 5 x 5 x 16 | 32 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Convolutional | 5 x 5 x 32 | 64 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Convolutional | 5 x 5 x 64 | 128 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Fully connected | 5 x 5 x 128 | 1024 |
| Fully connected | 1024 | 256 |
| Softmax | 256 | 60 |

Biểu diễn trực quan của mạng nơ-ron được sử dụng được đưa ra trong Hình 2.

* Lớp đầu tiên (Convolution #1) là lớp tích chập áp dụng 16 bộ lọc 5 x 5. Trên lớp này, chúng tôi áp dụng max pooling với bộ lọc hình dạng 2 x 2 với bước 2, chỉ định rằng các vùng được gộp không chồng lên nhau (Max-Pool #1). Điều này cũng làm giảm chiều rộng và chiều cao xuống 50 pixel mỗi vùng.
* Lớp tích chập thứ hai (Convolution #2) áp dụng 32 bộ lọc 5 x 5, tạo ra 32 bản đồ kích hoạt. Chúng tôi áp dụng trên lớp này cùng loại max pooling (Max-Pool #2) như trên lớp đầu tiên, hình dạng 2 x 2 và bước 2.
* Lớp tích chập thứ ba (Convolution #3) áp dụng 64 bộ lọc 5 x 5. Sau đây là một lớp nhóm tối đa khác (Max-Pool #3) có hình dạng 2 x 2 và bước 2.
* Lớp tích chập thứ tư (Convolution #4) áp dụng 128 bộ lọc 5 x 5 sau đó chúng ta áp dụng lớp nhóm tối đa cuối cùng (Max-Pool #4).



Hình 2: Biểu diễn đồ họa của mạng nơ-ron tích chập được sử dụng trong các thí nghiệm.

Chúng tôi trình bày một sơ đồ ngắn gọn về quy trình đào tạo:

| iterations = 75000 **read\_images**(images) **apply\_random\_hue\_saturation\_changes**(images) **apply\_random\_vertical\_horizontal\_flips**(images) **convert\_to\_hsv**(images) **add\_grayscale\_layer**(images) define\_network\_structure(images , network , training\_operation) for i in range(1, iterations): sess.run(training\_operation) |
| --- |

## 3.2 Đối tượng và mẫu nghiên cứu

Bảng 1: Số lượng hình ảnh cho mỗi loại trái cây. Có nhiều loại táo, mỗi loại được coi là một đối tượng riêng biệt. Chúng tôi không tìm thấy tên khoa học/tên phổ biến cho mỗi loại táo nên chúng tôi đã dán nhãn bằng chữ số (ví dụ: táo đỏ 1, táo đỏ 2, v.v.)

| Label | Số lượng hình ảnh huấn luyện | Số lượng hình ảnh thử nghiệm |
| --- | --- | --- |
| Apple Braeburn | 492 | 164 |
| Apple Golden 1 | 492 | 164 |
| Apple Golden 2 | 492 | 164 |
| Apple Golden 3 | 481 | 161 |
| Apple Granny Smith | 492 | 164 |
| Apple Red 1 | 492 | 164 |
| Apple Red 2 | 492 | 164 |
| Apple Red 3 | 429 | 144 |
| Apple Red Delicious | 490 | 166 |
| Apple Red Yellow 1 | 492 | 164 |
| Apple Red Yellow 2 | 672 | 219 |
| Apricot | 492 | 164 |
| Avocado | 427 | 143 |
| Avocado ripe | 491 | 166 |
| Banana | 490 | 166 |
| Banana Lady Finger | 450 | 152 |
| Banana Red | 490 | 166 |
| Cactus fruit | 490 | 166 |
| Cantaloupe 1 | 492 | 164 |

| Cantaloupe 2 | 492 | 164 |
| --- | --- | --- |
| Carambula | 490 | 166 |
| Cherry 1 | 492 | 164 |
| Cherry 2 | 738 | 246 |
| Cherry Rainier | 738 | 246 |
| Cherry Wax Black | 492 | 164 |
| Cherry Wax Red | 492 | 164 |
| Cherry Wax Yellow | 492 | 164 |
| Chestnut | 450 | 153 |
| Clementine | 490 | 166 |
| Cocos | 490 | 166 |
| Dates | 490 | 166 |
| Granadilla | 490 | 166 |
| Grape Blue | 984 | 328 |
| Grape Pink | 492 | 164 |
| Grape White | 490 | 166 |
| Grape White 2 | 490 | 166 |
| Grape White 3 | 492 | 164 |
| Grape White 4 | 471 | 158 |
| Grapefruit Pink | 490 | 166 |
| Grapefruit White | 492 | 164 |
| Guava | 490 | 166 |
| Huckleberry | 490 | 166 |
| Kaki | 490 | 166 |
| Kiwi | 466 | 156 |
| Kumquats | 490 | 166 |
| Lemon | 492 | 164 |
| Lemon Meyer | 490 | 166 |
| Limes | 490 | 166 |
| Lychee | 490 | 166 |
| Mandarine | 490 | 166 |
| Mango | 490 | 166 |
| Mangostan | 300 | 102 |
| Maracuja | 490 | 166 |

| Melon Piel de Sapo | 738 | 246 |
| --- | --- | --- |
| Mulberry | 492 | 164 |
| Nectarine | 492 | 164 |
| Orange | 479 | 160 |
| Papaya | 492 | 164 |
| Passion Fruit | 490 | 166 |
| Peach | 492 | 164 |
| Peach 2 | 738 | 246 |
| Peach Flat | 492 | 164 |
| Pear | 492 | 164 |
| Pear Abate | 490 | 166 |
| Pear Monster | 490 | 166 |
| Pear Williams | 490 | 166 |
| Pepino | 490 | 166 |
| Physalis | 492 | 164 |
| Physalis with Husk | 492 | 164 |
| Pineapple | 490 | 166 |
| Pineapple Mini | 493 | 163 |
| Pitahaya Red | 490 | 166 |
| Plum | 447 | 151 |
| Pomegranate | 492 | 164 |
| Quince | 490 | 166 |
| Rambutan | 492 | 164 |
| Raspberry | 490 | 166 |
| Redcurrant | 492 | 164 |
| Salak | 490 | 162 |
| Strawberry | 492 | 164 |
| Strawberry Wedge | 738 | 246 |
| Tamarillo | 490 | 166 |
| Tangelo | 490 | 166 |
| Tomato 1 | 738 | 246 |
| Tomato 2 | 672 | 225 |
| Tomato 3 | 738 | 246 |
| Tomato 4 | 479 | 160 |

## 3.3. Cách thu thập dữ liệu

Trong phần này, chúng tôi mô tả cách tạo bộ dữ liệu và nội dung của bộ dữ liệu. Hình ảnh được thu được bằng cách quay phim các loại trái cây trong khi chúng được một động cơ quay và sau đó trích xuất các khung hình.

Các loại trái cây được đặt trong trục của một động cơ tốc độ thấp (3 vòng/phút) và một bộ phim ngắn 20 giây đã được ghi lại. Đằng sau các loại trái cây, chúng tôi đặt một tờ giấy trắng làm nền.



Hình 1: Bên trái: hình ảnh gốc. Lưu ý phần nền và trục động cơ. Bên phải: quả sau khi xóa phần nền và sau khi được thu nhỏ xuống 100x100 pixel.

## 3.4 Tiền xử lý dữ liệu.

Do điều kiện ánh sáng thay đổi, nền không đồng nhất và chúng tôi đã viết một thuật toán chuyên dụng để trích xuất trái cây từ nền. Thuật toán này thuộc loại tô màu tràn: chúng tôi bắt đầu từ mỗi cạnh của hình ảnh và đánh dấu tất cả các pixel ở đó, sau đó chúng tôi đánh dấu tất cả các pixel được tìm thấy trong vùng lân cận của các pixel đã đánh dấu mà khoảng cách giữa các màu nhỏ hơn một giá trị quy định. chúng tôi lặp lại bước trước cho đến khi không thể đánh dấu thêm pixel nào nữa.

Tất cả các pixel được đánh dấu đều được coi là nền (sau đó được tô màu trắng) và các pixel còn lại được coi là thuộc về đối tượng. Giá trị tối đa cho khoảng cách giữa 2 pixel lân cận là tham số của thuật toán và được thiết lập (bằng cách thử và sai) cho mỗi bộ phim.

Các loại trái cây được chia tỷ lệ để phù hợp với hình ảnh 100x100 pixel. Các tập dữ liệu khác (như MNIST) sử dụng hình ảnh 28x28, nhưng chúng tôi cảm thấy rằng kích thước nhỏ sẽ gây bất lợi khi bạn có các đối tượng quá giống nhau (một quả anh đào đỏ trông rất giống một quả táo đỏ trong các hình ảnh nhỏ). Kế hoạch trong tương lai của chúng tôi là làm việc với các hình ảnh thậm chí còn lớn hơn, nhưng điều này sẽ đòi hỏi thời gian đào tạo lâu hơn nhiều.

## 3.5 Thư viện TensorFlow.

Với mục đích triển khai, đào tạo và thử nghiệm mạng được mô tả trong bài báo này, chúng tôi đã sử dụng thư viện TensorFlow [32]. Đây là một khuôn khổ nguồn mở cho máy học do Google tạo ra để tính toán số bằng biểu đồ luồng dữ liệu. Các nút trong biểu đồ biểu diễn các phép toán, trong khi các cạnh biểu đồ biểu diễn các mảng dữ liệu đa chiều được gọi là tenxơ.

Các thành phần chính trong hệ thống TensorFlow là máy khách, sử dụng giao diện Phiên để giao tiếp với máy chủ và một hoặc nhiều quy trình công nhân, trong đó mỗi quy trình công nhân chịu trách nhiệm phân xử quyền truy cập vào một hoặc nhiều thiết bị tính toán (như lõi CPU hoặc thẻ GPU) và thực thi các nút biểu đồ trên các thiết bị đó theo hướng dẫn của máy chủ.

Chúng tôi sẽ trình bày các phương pháp và kiểu dữ liệu được sử dụng quan trọng nhất từ TensorFlow cùng với mô tả ngắn gọn cho từng phương pháp. Một lớp tích chập được định nghĩa như sau:

| conv2d( input , filter , strides , padding , use\_cudnn\_on\_gpu=True , data\_format='NHWC', dilations=[1, 1, 1, 1], name=None ) |
| --- |

Tính toán tích chập 2 chiều cho các tenxơ lọc và đầu vào 4 chiều. Cho một tenxơ đầu vào có hình dạng [lô, theo chiều cao, theo chiều rộng, theo kênh] và một tenxơ hạt nhân có hình dạng [chiều cao bộ lọc, chiều rộng bộ lọc, theo kênh, theo kênh ra], thao tác này thực hiện như sau:

* Làm phẳng bộ lọc thành ma trận 2 chiều có hình dạng [chiều cao bộ lọc \* chiều rộng bộ lọc \* theo kênh, kênh ra].
* Trích xuất các bản vá hình ảnh từ tenxơ đầu vào để tạo thành một tenxơ ảo có hình dạng [lô, chiều cao bộ lọc, chiều rộng bộ lọc, chiều cao bộ lọc \* chiều rộng bộ lọc \* theo kênh].
* Đối với mỗi bản vá, nhân ma trận bộ lọc và vectơ bản vá hình ảnh theo chiều phải.

| tf.nn.max\_pool( value , ksize , strides , padding , data\_format='NHWC', name=None ) |
| --- |

Thực hiện thao tác gom nhóm tối đa trên đầu vào. Các tham số ksize và strides có thể là các bộ hoặc danh sách các bộ gồm 4 phần tử. Ksize biểu thị kích thước của cửa sổ cho mỗi chiều của tenxơ đầu vào và strides biểu thị bước tiến của cửa sổ trượt cho mỗi chiều của tenxơ đầu vào. Tham số đệm có thể là ‘’VALID’‘ hoặc ‘’SAME’‘.

| tf.nn.relu( features , name=None ) |
| --- |

Tính toán phép toán tuyến tính đã chỉnh lưu - max(features, 0). Features là một tenxơ.

| tf.nn.dropout( x, keep\_prob , noise\_shape=None , seed=None , name=None ) |
| --- |

Áp dụng dropout trên đầu vào x với xác suất giữ prob. Điều này có nghĩa là đối với mỗi giá trị trong x, phương pháp đưa ra giá trị được chia tỷ lệ theo 1 / giữ prob với xác suất giữ prob hoặc 0. Việc chia tỷ lệ được thực hiện theo thứ tự để bảo toàn tổng các phần tử. Tham số hình dạng nhiễu xác định nhóm giá trị nào được giữ lại hoặc loại bỏ cùng nhau. Ví dụ: giá trị [k, 1, 1, n] cho hình dạng nhiễu, với x có hình dạng [k, l, m, n], có nghĩa là mỗi hàng và cột sẽ được giữ lại hoặc loại bỏ cùng nhau, trong khi các thành phần lô và kênh sẽ được giữ lại hoặc loại bỏ riêng biệt.

# Chương 4. Thực nghiệm và Thảo luận

## 4.1. Thí nghiệm số.

Bộ dữ liệu được chia thành 2 phần: tập huấn luyện - bao gồm 46371 hình ảnh về trái cây và tập kiểm tra - bao gồm 15563 hình ảnh.

Dữ liệu được đóng gói thành tệp TFRecords (dành riêng cho TensorFlow). Đây là tệp nhị phân chứa bộ đệm giao thức có bản đồ đặc điểm. Trong bản đồ này có thể lưu trữ thông tin như chiều cao, chiều rộng, chiều sâu của hình ảnh và thậm chí cả hình ảnh thô. Sử dụng các tệp này, chúng ta có thể tạo hàng đợi để cung cấp dữ liệu cho mạng nơ-ron.

Chúng tôi đã chạy nhiều kịch bản trong đó mạng nơ-ron được đào tạo bằng cách sử dụng các mức tăng cường dữ liệu và tiền xử lý khác nhau:

* chuyển đổi hình ảnh RGB đầu vào thành thang độ xám.
* giữ nguyên hình ảnh đầu vào trong không gian màu RGB.
* chuyển đổi hình ảnh RGB đầu vào thành không gian màu HSV.
* chuyển đổi hình ảnh RGB đầu vào thành không gian màu HSV và thang độ xám và hợp nhất chúng.
* áp dụng các thay đổi ngẫu nhiên về sắc độ và độ bão hòa trên hình ảnh RGB đầu vào, lật ngẫu nhiên chúng theo chiều ngang và chiều dọc, sau đó chuyển đổi chúng thành không gian màu HSV và thang độ xám rồi hợp nhất chúng.

## 4.2 Đào tạo mạng lưới thần kinh

Đối với mỗi kịch bản, chúng tôi sử dụng mạng nơ-ron đã mô tả trước đó đã được đào tạo qua 75000 lần lặp với các lô 60 hình ảnh được chọn ngẫu nhiên từ tập đào tạo. Cứ sau 50 bước, chúng tôi tính toán độ chính xác bằng cách sử dụng xác thực chéo. Để thử nghiệm, chúng tôi chạy mạng đã được đào tạo trên tập thử nghiệm. Kết quả cho từng trường hợp được trình bày trong Bảng 3.

Bảng 3: Kết quả đào tạo mạng nơ-ron trên tập dữ liệu fruits-360.

| Scenario | Accuracy on training set | Accuracy on test set |
| --- | --- | --- |
| Grayscale | 99.96% | 94.24% |
| RGB | 99.23% | 93.47% |
| HSV | 99.98% | 97.01% |
| HSV + Grayscale | 99.78% | 95.71% |
| HSV + Grayscale + hue/saturation  change + flips | 99.86% | 97.04% |

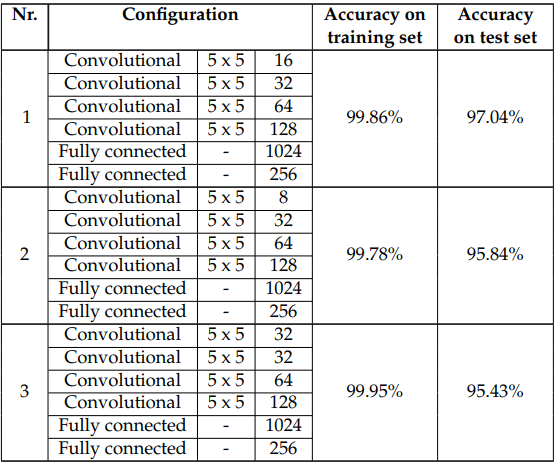
Như được phản ánh trong Bảng 3, kết quả tốt nhất thu được bằng cách áp dụng tăng cường dữ liệu và chuyển đổi hình ảnh RGB sang không gian màu HSV mà biểu diễn thang độ xám đã được thêm vào. Điều này rất trực quan vì trong kịch bản này, chúng tôi đính kèm lượng thông tin lớn nhất vào đầu vào, do đó mạng có thể học nhiều tính năng để phân loại hình ảnh.

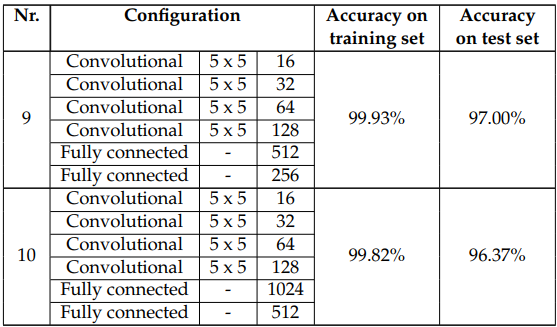
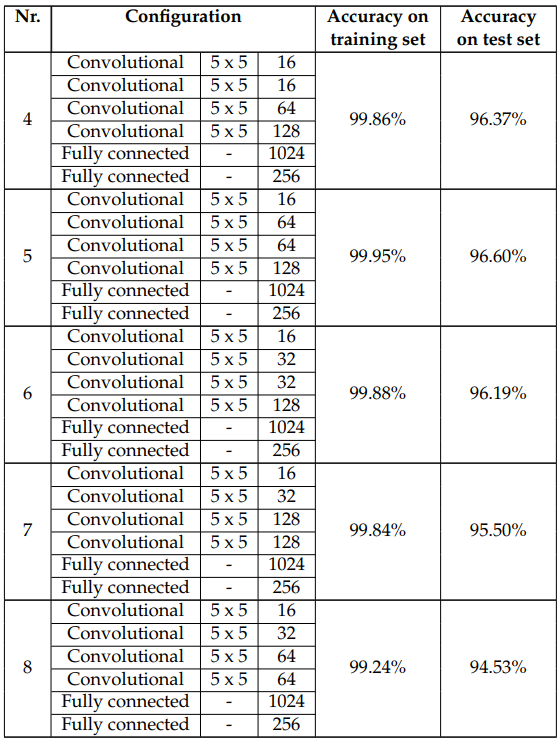
Điều quan trọng cần lưu ý là việc đào tạo các hình ảnh thang độ xám chỉ mang lại kết quả tốt nhất trên tập huấn luyện nhưng lại cho kết quả rất yếu trên tập kiểm tra. Chúng tôi đã điều tra vấn đề này và phát hiện ra rằng rất nhiều hình ảnh chứa táo được phân loại không chính xác trên tập kiểm tra. Để tiếp tục điều tra vấn đề này, chúng tôi đã chạy một vòng đào tạo và thử nghiệm chỉ trên các lớp hình ảnh táo. Kết quả tương tự, với độ chính xác cao trên dữ liệu huấn luyện, nhưng độ chính xác thấp trên dữ liệu kiểm tra. Chúng tôi cho rằng điều này là do quá trình khớp, vì các hình ảnh thang độ xám mất quá nhiều tính năng nên mạng không học đúng cách để phân loại các hình ảnh.

### 

## 4.3. Đào tạo các cấu hình mạng khác nhau.

Để xác định cấu hình mạng tốt nhất để phân loại hình ảnh trong tập dữ liệu của chúng tôi, chúng tôi đã lấy nhiều cấu hình, sử dụng tập huấn luyện để huấn luyện chúng và sau đó tính toán độ chính xác của chúng trên tập kiểm tra và tập huấn luyện. Trong Bảng 4, chúng tôi trình bày kết quả.

Bảng 4: Kết quả huấn luyện các cấu hình mạng khác nhau trên tập dữ liệu fruits-360.



Từ Bảng 4, chúng ta có thể thấy rằng hiệu suất tốt nhất trên tập kiểm tra thu được bởi cấu hình số 1, tuy nhiên, cấu hình này không thu được độ chính xác tốt nhất trên tập huấn luyện. Cấu hình 5 và 3 thu được độ chính xác tốt nhất trên tập huấn luyện (99,95%), nhưng hiệu suất trên tập kiểm tra tụt hậu một chút. Đối với cấu hình 3, hiện tượng này rõ ràng hơn nhiều, vì sự khác biệt giữa độ chính xác của tập huấn luyện và kiểm tra là 4,52%. Tương tự, cấu hình 8 có sự khác biệt lớn về hiệu suất giữa tập huấn luyện và tập kiểm tra. Đây là kết quả của việc mô hình quá khớp với dữ liệu huấn luyện và không khái quát hóa đúng cách cho các hình ảnh khác. Hiệu suất tốt thứ hai trên tập huấn luyện thu được bởi cấu hình 9, cấu hình này cũng thu được độ chính xác tốt thứ hai trên tập kiểm tra.

## 4.4 Sai số thực nghiệm.

Sự tiến triển của độ chính xác trong quá trình đào tạo được đưa ra trong Hình 3. Có thể thấy rằng quá trình đào tạo được cải thiện nhanh chóng trong 1000 lần lặp đầu tiên (độ chính xác trở nên lớn hơn 90%) và sau đó được cải thiện rất chậm trong 74000 lần lặp tiếp theo.

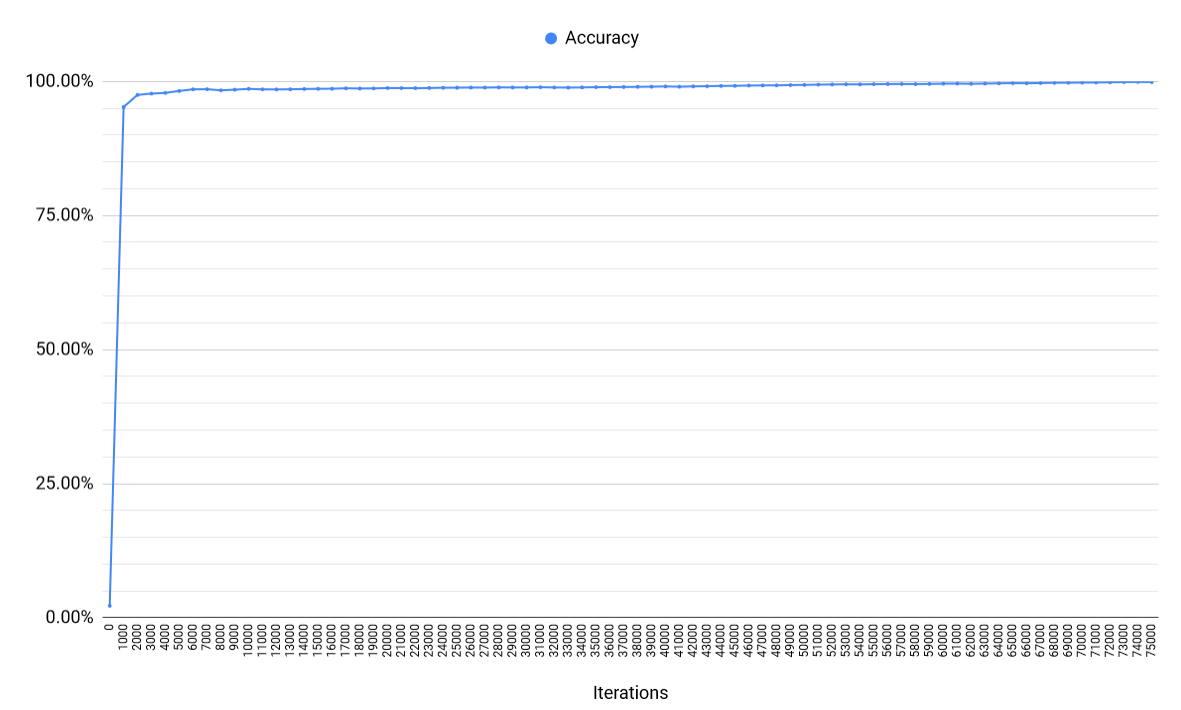
Một số hình ảnh được phân loại không chính xác được đưa ra trong Bảng 5.

Bảng 5: Một số hình ảnh được phân loại không chính xác. Ở trên cùng, chúng ta có lớp trái cây chính xác và ở dưới cùng, chúng ta có lớp (và xác suất liên quan) được mạng chỉ định.

| Apple Golden 2    Apple Golden 3  96.54% | Apple Golden 3    Granny Smith  (Apple)  95.22% | Braeburn(Apple)    Apple Red 2  97.71% | Peach    Apple Red  Yellow  97.85% |
| --- | --- | --- | --- |
| Pomegranate    Nectarine  94.64% | Peach    Apple Red 1  97.87% | Pear    Apple Golden 2  98.73% | Pomegranate    Braeburn(Apple)  97.21% |

# Chương 5. Kết luận và Hướng phát triển

### 5.1. Tóm tắt kết quả đạt được và hướng nghiên cứu tương lai.

Chúng tôi đã mô tả một cơ sở dữ liệu hình ảnh mới và phức tạp với các loại trái cây. Chúng tôi cũng đã thực hiện một số thí nghiệm số bằng cách sử dụng thư viện TensorFlow để phân loại hình ảnh theo nội dung của chúng.  


Hình 3: Sự cải thiện độ chính xác qua 75000 lần lặp lại đào tạo.

Theo quan điểm của chúng tôi, một trong những mục tiêu chính trong tương lai là cải thiện độ chính xác của mạng nơ-ron. Điều này liên quan đến việc thử nghiệm thêm với cấu trúc của mạng. Nhiều điều chỉnh và thay đổi khác nhau đối với bất kỳ lớp nào cũng như việc đưa vào các lớp mới có thể cung cấp kết quả hoàn toàn khác nhau. Một lựa chọn khác là thay thế tất cả các lớp bằng các lớp tích chập . Điều này đã được chứng minh là cung cấp một số cải tiến so với các mạng có các lớp được kết nối đầy đủ trong cấu trúc của chúng. Một hệ quả của việc thay thế tất cả các lớp bằng các lớp tích chập là sẽ có sự gia tăng về số lượng tham số cho mạng [29]. Một khả năng khác là thay thế các đơn vị tuyến tính chỉnh lưu bằng các đơn vị tuyến tính mũ. Theo bài báo [8], điều này làm giảm độ phức tạp của tính toán và tăng đáng kể

hiệu suất tổng quát hóa tốt hơn so với các đơn vị tuyến tính chỉnh lưu trên các mạng có hơn 5 lớp. Chúng tôi muốn thử các phương pháp này và cũng cố gắng tìm các cấu hình mới cung cấp kết quả thú vị.

Trong tương lai gần, chúng tôi có kế hoạch tạo một ứng dụng di động chụp ảnh trái cây và dán nhãn cho chúng theo đó.

.

### 5.2. Câu trả lời cho câu hỏi nghiên cứu

*Liệu có thể xây dựng một hệ thống tự động xử lý và nhận dạng hình ảnh trái cây quay bằng thiết bị đơn giản và nền trắng không đồng đều hay không?*

* Với việc áp dụng thuật toán định lượng (tách nền bằng flood fill) và mô hình CNN tự xây dựng, bài toán nhận dạng trái cây từ video thu được đã được giải quyết một cách hiệu quả.
* Quá trình tách nền hoạt động tốt trong hầu hết các trường hợp và mô hình học sâu đã khai thác tốt dữ liệu đầu vào sạch.

# TÀI LIỆU THAM KHẢO

[1] Bargoti, S., and Underwood, J. Deep fruit detection in orchards. In 2017 IEEE International Conference on Robotics and Automation (ICRA) (May 2017), pp. 3626–3633.

[2] Barth, R., IJsselmuiden, J., Hemming, J., and Henten, E. V. Data synthesis methods for semantic segmentation in agriculture: A capsicum annuum dataset. Computers and Electronics in Agriculture 144 (2018), 284 – 296.

[3] Chan, T. F., and Vese, L. A. Active contours without edges. IEEE Transactions on Image Processing 10, 2 (Feb 2001), 266–277.

[4] Cheng, H., Damerow, L., Sun, Y., and Blanke,M. Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. Journal of Imaging 3, 1 (2017).

[5] Cires¸an, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2 (USA, 2012), NIPS’12, Curran Associates Inc., pp. 2843–2851.

[6] Cires¸an, D. C., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J. Flexible, high performance convolutional neural networks for image classification. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two (2011), IJCAI’11, AAAI Press, pp. 1237–1242.

[7] Ciresan, D. C., Meier, U., and Schmidhuber, J. Multi-column deep neural networks for image classification. CoRR abs/1202.2745 (2012).

[8] Clevert, D., Unterthiner, T., and Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). CoRR abs/1511.07289 (2015).

[9] Hannun, A. Y., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., and Ng, A. Y. Deep speech: Scaling up end-to-end speech recognition. CoRR abs/1412.5567 (2014).

[10] Hemming, J., Ruizendaal, J., Hofstee, J. W., and van Henten, E. J. Fruit detectability analysis for different camera positions in sweet-pepper. Sensors 14, 4 (2014), 6032–6044.

[11] Kapach, K., Barnea, E., Mairon, R., Edan, Y., and Ben-Shahar, O. Computer vision for fruit harvesting robots – state of the art and challenges ahead. Int. J. Comput. Vision Robot. 3, 1/2 (Apr. 2012), 4–34.

[12] Krizhevsky, A., Nair, V., and Hinton, G. The cifar dataset. [Online; accessed 27.10.2018].

[13] LeCun, Y., Cortes, C., and Burges, C. J. The mnist database of handwritten digits. [Online; accessed 27.10.2018].

[14] Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th Annual International Conference on Machine Learning (New York, NY, USA, 2009), ICML ’09, ACM, pp. 609–616.

[15] Li, D., Zhao, H., Zhao, X., Gao, Q., and Xu, L. Cucumber detection based on texture and color in greenhouse. International Journal of Pattern Recognition and Artificial Intelligence 31 (01 2017).

[16] Liang, M., and Hu, X. Recurrent convolutional neural network for object recognition. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015), pp. 3367–3375.

[17] Mumford, D., and Shah, J. Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics 42, 5 (1989), 577–685.

[18] Muresan, H., and Oltean, M. Fruits 360 dataset on github. [Online; accessed 27.10.2018].

[19] Muresan, H., and Oltean, M. Fruits 360 dataset on kaggle. [Online; accessed 27.10.2018].

[20] Ninawe, P., and Pandey, M. S. A completion on fruit recognition system using k-nearest neighbors algorithm. In International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) (2014), vol. 3.

[21] O’Boyle, B., and Hall, C. What is google lens and how do you use it? [Online; accessed 05.05.2018]. )2

[22] Puttemans, S., Vanbrabant, Y., Tits, L., and Goedem, T. Automated visual fruit detection for harvest estimation and robotic harvesting. In 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA) (Dec 2016), pp. 1–6.

[23] Rahnemoonfar, M., and Sheppard, C. Deep count: Fruit counting based on deep simulated learning. Sensors 17, 4 (2017).

[24] Ren, S., He, K., Girshick, R. B., and Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. CoRR abs/1506.01497 (2015).

[25] Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and McCool, C. Deepfruits: A fruit detection system using deep neural networks. Sensors 16, 8 (2016).

[26] Schmidhuber, J. Deep learning in neural networks: An overview. CoRR abs/1404.7828 (2014).

[27] Selvaraj, A., Shebiah, N., Nidhyananthan, S., and Ganesan, L. Fruit recognition using color and texture features. Journal of Emerging Trends in Computing and Information Sciences 1 (10 2010), 90–94.

[28] Song, Y., Glasbey, C., Horgan, G., Polder, G., Dieleman, J., and van der Heijden, G. Automatic fruit recognition and counting from multiple images. Biosystems Engineering 118 (2014), 203 – 215.

[29] Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. A. Striving for simplicity: The all convolutional net. CoRR abs/1412.6806 (2014).

[30] Srivastava, R. K., Greff, K., and Schmidhuber, J. Training very deep networks. CoRR abs/1507.06228 (2015).

[31] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. CoRR abs/1409.4842 (2014).

[32] TensorFlow. Tensorflow. [Online; accessed 05.05.2018].

[33] Wikipedia. Convolution in mathematics. [Online; accessed 05.05.2018].

[34] Wikipedia. Deep learning article on wikipedia. [Online; accessed 05.05.2018].

[35] Wikipedia. Google lens on wikipedia. [Online; accessed 05.05.2018].

[36] Xiong, J., Liu, Z., Lin, R., Bu, R., He, Z., Yang, Z., and Liang, C. Green grape detection and picking-point calculation in a night-time natural environment using a charge-coupled device (ccd) vision sensor with artificial illumination. Sensors 18, 4 (2018).

[37] Zawbaa, H., Abbass, M., Hazman, M., and Hassanien, A. E. Automatic fruit image recognition system based on shape and color features. Communications in Computer and Information Science 488 (11 2014), 278– 290.

# PHỤ LỤC