Shape, square

Description automatically generated

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

**KHOA CÔNG NGHỆ THÔNG TIN**

-----🙞🙜🕮🙞🙜-----

Logo, company name

Description automatically generated

**LUẬN VĂN TỐT NGHIỆP**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

**Đề tài**

**NHẬN DẠNG HOA**

**SỬ DỤNG MÔ HÌNH HỌC SÂU**

FLOWER RECOGNITION

USING DEEP LEARNING MODELS

**Sinh viên: Nguyễn Thị Anh Thư**

**Mã số: B1809301**

**Khóa: 44**

***Cần Thơ, 12/2022***

Shape, square

Description automatically generated

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

**KHOA CÔNG NGHỆ THÔNG TIN**

-----🙞🙜🕮🙞🙜-----

Logo, company name

Description automatically generated

**LUẬN VĂN TỐT NGHIỆP**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

**Đề tài**

**NHẬN DẠNG HOA**

**SỬ DỤNG MÔ HÌNH HỌC SÂU**

FLOWER RECOGNITION

USING DEEP LEARNING MODELS

**Giáo viên hướng dẫn: Sinh viên thực hiện:**

**TS. Phạm Thế Phi Nguyễn Thị Anh Thư**

**Mã số: B1809301**

**Khóa : 44**

***Cần Thơ, 12/2022***

TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

**KHOA CÔNG NGHỆ THÔNG TIN**

**XÁC NHẬN CHỈNH SỬA LUẬN VĂN**

**THEO YÊU CẦU CỦA HỘI ĐỒNG**

Tên luận văn (tiếng Việt và tiếng Anh): Nhận dạng hoa sử dụng mô hình học sâu (Flower recognition using deep learning models)

Họ tên sinh viên: Nguyễn Thị Anh Thư MASV: B1809301

Mã lớp: DI18V7A2

Đã báo cáo tại hội đồng ngành: Công nghệ thông tin

Ngày báo cáo: 13/12/2022

Hội đồng báo cáo gồm:

1. T.S. Lâm Nhựt Khang Chủ tịch hội đồng
2. T.S. Bùi Võ Quốc Bảo Thành viên
3. T.S. Phạm Thế Phi Thư ký

Luận văn đã được chỉnh sửa theo góp ý của Hội đồng.

*Cần Thơ, ngày ….. tháng …… năm 20…*

**Giáo viên hướng dẫn**

*(Ký và ghi họ tên)*

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

🙠 🕮 🙢

*Cần Thơ, ngày ...... tháng ...... năm 20...*

**Giáo viên hướng dẫn**

*(Ký và ghi rõ họ tên)*

*........................................................*

**LỜI CAM ĐOAN**

Em xin được cam đoan luận văn kèm với đề tài có tên là “Nhận dạng hoa sử dụng mô hình học sâu” được triển khai thực hiện hoàn toàn dựa trên sự học hỏi, tìm hiểu và kết quả nghiên cứu của em dưới sự hướng dẫn của thầy Phạm Thế Phi.

Các nguồn tài liệu tham khảo đã được trình bày rõ trong quyển báo cáo thuộc danh mục tài liệu tham khảo.

Cần Thơ, ngày 01 tháng 12 năm 2022

Sinh viên thực hiện

Nguyễn Thị Anh Thư

**LỜI CẢM ƠN**

Để có được luận văn này, em xin gửi lời cảm ơn chân thành đến Ts. Phạm Thế Phi, người đã dành thời gian ra để tận tình giúp đỡ em. Trong suốt quá trình thực hiện đề tài, nhờ sự hướng dẫn tận tình, tâm huyết của thầy cũng như những kiến thức quý báo mà thầy truyền đạt đã giúp cho bài luận văn này được hoàn thành một cách tốt nhất.

Em cũng xin gửi lời cảm ơn sâu sắc đến các Thầy Cô trường Đại học Cần Thơ, đặt biệt là các Thầy Cô trường Công nghệ Thông tin và Truyền thông, những người đã tận tình truyền đạt kiến thức quý giá cho chúng em trong thời gian qua.

Ngoài ra, em cũng xin cảm ơn gia đình và bạn bè đã luôn luôn động viên và tạo mọi điều kiện cho em trong suốt quá trình học tập. Em xin cảm ơn các anh, chị đã giúp đỡ, góp ý và chỉ bảo tận tình để em có thể hoàn thành bài luận văn này.

Tuy có nhiều cố gắng trong quá trình thực hiện luận văn, nhưng không thế tránh khỏi những thiếu sót. Em rất mong nhận được sự góp ý, chỉ bảo của các Thầy Cô để bài luận văn trở nên hoàn thiện hơn.

Em xin chân thành cảm ơn!

Cần Thơ, ngày 01 tháng 12 năm 2022

Sinh viên thực hiện

Nguyễn Thị Anh Thư

# **MỤC LỤC**

[**MỤC LỤC** 1](#_Toc124187252)

[**DANH MỤC HÌNH** 3](#_Toc124187253)

[**DANH MỤC BẢNG** 4](#_Toc124187254)

[**DANH MỤC TỪ VIẾT TẮT** 5](#_Toc124187255)

[**TÓM TẮT** 6](#_Toc124187256)

[**ABSTRACT** 7](#_Toc124187257)

[**CHƯƠNG 1. GIỚI THIỆU** 8](#_Toc124187258)

[**1.1** **Đặt vấn đề** 8](#_Toc124187259)

[**1.2** **Lịch sử giải quyết vấn đề** 9](#_Toc124187260)

[**1.3** **Mục tiêu đề tài** 9](#_Toc124187261)

[**1.4** **Đối tượng và phạm vi nghiên cứu** 10](#_Toc124187262)

[**1.5** **Phương pháp nghiên cứu** 10](#_Toc124187263)

[**1.6** **Kết quả đạt được** 11](#_Toc124187264)

[**1.7** **Bố cục luận văn** 11](#_Toc124187265)

[**CHƯƠNG 2. CƠ SỞ LÝ THUYẾT** 12](#_Toc124187266)

[**2.1** **Thị giác máy tính** 12](#_Toc124187267)

[**2.2** **Mô hình học sâu** 12](#_Toc124187268)

[**2.3** **Mô hình mạng nơ-ron tích chập** 13](#_Toc124187269)

[**2.4** **Các phương pháp tiếp cận trong nhận dạng hoa** 17](#_Toc124187270)

[**2.4.1.** **Mô hình Efficientnet-B5** 17](#_Toc124187276)

[**2.4.2.** **Mô hình Resnet-50** 17](#_Toc124187277)

[**2.4.3.** **Mô hình Inception-V3** 19](#_Toc124187278)

[**CHƯƠNG 3. PHƯƠNG PHÁP THỰC HIỆN** 22](#_Toc124187279)

[**3.1** **Tổng quan** 22](#_Toc124187280)

[**3.2** **Cài đặt giải thuật** 23](#_Toc124187281)

[**3.2.1.** **Thu thập dữ liệu** 23](#_Toc124187282)

[**3.2.2.** **Huấn luyện mô hình** 25](#_Toc124187283)

[**3.2.3.** **Thiết kế giao diện** 26](#_Toc124187284)

[**CHƯƠNG 4. KẾT QUẢ THỰC NGHIỆM** 28](#_Toc124187285)

[**4. 1** **Kết quả kiểm tra** 28](#_Toc124187286)

[**4. 2** **Đánh giá kết quả** 31](#_Toc124187287)

[**CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN** 32](#_Toc124187288)

[**5.1** **Kết quả đạt được** 32](#_Toc124187289)

[**5.2** **Hướng phát triển** 32](#_Toc124187290)

[**TÀI LIỆU THAM KHẢO** 33](#_Toc124187291)

# **DANH MỤC HÌNH**

[Hình 2.1. Cấu trúc mạng nơ-ron tích chập đơn giản 14](#_Toc124187380)

[Hình 2.2. Minh họa phép tích chập 15](#_Toc124187381)

[Hình 2.3. Hàm kích hoạt phi tuyến ReLU 16](#_Toc124187382)

[Hình 2.4. Minh họa phép Max Pooling trên ma trận 4x4 16](#_Toc124187383)

[Hình 2.5. Kiến trúc ResNet bao gồm 2 khối đặt trưng 17](#_Toc124187384)

[Hình 2.6. Cộng trực tiếp đầu vào của khối với nhánh còn lại trong khối Identity block 18](#_Toc124187385)

[Hình 2.7. Kiến trúc tóm tắt của mạng Resnet-50 19](#_Toc124187386)

[Hình 2.8. Kiến trúc GoogleNet – Inception version 3 20](#_Toc124187387)

[Hình 3.1. Sơ đồ các bước thực hiện 22](#_Toc124187388)

[Hình 3.2. Hình ảnh ví dụ của Hoa Cẩm Chướng 23](#_Toc124187389)

[Hình 3.3. Hình ảnh ví dụ của Hoa Đồng Tiền 24](#_Toc124187390)

[Hình 3.4. Giao diện của ứng dụng web khi load trang 26](#_Toc124187391)

[Hình 3.5. Giao diện kết quả trả về sau khi upload hình ảnh lên web 27](#_Toc124187392)

# **DANH MỤC BẢNG**

[Bảng 3.1. Tập dữ liệu hình ảnh 24](#_Toc124187398)

[Bảng 3.2. Độ chính xác của từng mô hình 25](#_Toc124187399)

[Bảng 4.1. Bảng Confusion maxtix cho lớp Hoa Dâm Bụt 29](#_Toc124187404)

[Bảng 4.2. Thống kê số lượng hình ảnh sau kiểm thử 29](#_Toc124187405)

[Bảng 4.3. mAP mô hình Inception-V3 30](#_Toc124187406)

[Bảng 4.4. Giá trị trung bình mAP của mô hình Inception-V3 31](#_Toc124187407)

# **DANH MỤC TỪ VIẾT TẮT**

|  |  |
| --- | --- |
| **Từ viết tắt** | **Viết đầy đủ** |
| AI | Artificial Intelligence |
| API | Application Programming Interface |
| CNN | Convolutional Neural Network |
| FN | False Negative |
| FP | False positive |
| mAP | mean Average Precision |
| ReLU | Rectified Linear Unit |
| SIANN | Shift Invariant Artificial Neural Network |
| TP | True positive |

# **TÓM TẮT**

Hiện nay hoa xuất hiện ở khắp nơi, đặt biệt là các địa điểm thu hút nhiều khách du lịch bởi vẻ đẹp và sự đa dạng của chúng. Việc nhận biết hoa cũng là một vấn đề hết sức nan giải đối với nhiều người vì hoa không chỉ có môt hoặc hai loại mà là rất nhiều. Cùng với sự phát triển của khoa học công nghệ, các thuật toán học máy như mạng lưới thần kinh tích chập ra đời dùng để xác định các loài hoa góp phần cho sự phát triển của nhiều ngành khoa học trong tương lai. Trong bài luận văn này, sử dụng cách tiếp cận dựa trên mô hình mạng nơ-ron tích chập (CNN) cho bài toán “Nhận dạng hoa sử dụng mô hình học sâu”. Tập dữ liệu gồm 1433 hình ảnh của 25 loài hoa khác nhau, hình ảnh được đưa về dạng “.jpg” để huấn luyện mô hình. Kiểm tra và đánh giá mô hình với 286 ảnh của 25 loài hoa tương ứng. Kết quả mô hình đạt được 86% độ chính xác. Xây dựng ứng dụng web nhận dạng với mô hình Inception-V3, giao diện của ứng dụng đơn giản, dễ sử dụng.

# **ABSTRACT**

Currently, flowers appear everywhere, especially in places that attract many tourists because of their beauty and diversity. The identification of flowers is also a very difficult problem for many people because flowers are not only one or two types but many. Along with the development of science and technology, machine learning algorithms such as convolutional neural networks were born to identify flowers, contributing to the development of many sciences in the future. In this thesis, an approach based on convolutional neural network (CNN) model is used for the problem of "Flower recognition using deep learning models". The data set includes 1433 images of 25 different flowers, the images are converted to ".jpg" format to train the model. Test and evaluate the model with 286 photos of 25 respective flower species. The model results achieved 86% accuracy. Build identity web application with Inception-V3 model, the interface of the application is simple and easy to use.

# 

# **CHƯƠNG 1. GIỚI THIỆU**

Trong thời đại phát triển như hiện nay, việc ứng dụng các tiến bộ của khoa học công nghệ đã trở nên phổ biến với mọi người, các hệ thống nhận dạng hình ảnh cũng dần trở thành một phần không thể thiếu trong nhiều lĩnh vực. Luận văn này trình bày các vấn đề về hệ thống “Nhận dạng hoa”. Chương 1 sẽ trình bày các vấn đề tổng quan về nhận dạng hoa, lịch sử giải quyết vấn đề, mục tiêu đề tài, đối tượng và phạm vi nghiên cứu, phương pháp nghiên cứu và kết quả đạt được.

* 1. **Đặt vấn đề**

Hoa xuất hiện ở khắp mọi nơi trên thế giới, chúng có nhiều công dụng khác nhau đối với con người và cả các loài sinh vật khác. Một số loài được dùng để trang trí nhà cửa, làm quà tặng trong các ngày lễ đặt biệt, một số được dùng cho việc thu hút khách du lịch ở nhiều nơi. Ngoài ra, chúng còn được dùng làm thức ăn cho các loài động vật cũng như làm thuốc chữa bệnh cho con người. Tuy nhiên, không phải ai cũng có thể biết được hết các loài hoa để đưa ra sự lựa chọn phù hợp. Do đó, việc hiểu biết về hoa là điều hết sức cần thiết để xác định được tên hoa là gì, hoa đó thuộc loại nào, công dụng ra sao và cách chăm sóc cho từng loài hoa như thế nào. Điều này sẽ giúp ích rất nhiều cho cả người sử dụng, người kinh doanh hoa cảnh, cho ngành du lịch và ngành công nghiệp dược liệu của thế giới.

Việc phân loại hoa là một việc quan trọng, nó đã được nghiên cứu và được tiếp cận bằng nhiều cách khác nhau. Nhiều người nhận biết và phân biệt hoa bằng những dấu hiệu đặt trưng về cấu tạo, hình dáng và màu sắt bên ngoài của từng loại hoa đòi hỏi độ hiểu biết và sự nhạy bén của từng người phải cao. Trước đây cũng đã có nhiều cách nhận dạng hoa bằng cách sử dụng kỹ thuật thủ công và đã được áp dụng nhưng cho kết quả chính xác không cao. Do đó, việc nhận dạng và phân loại hoa còn khá khó khăn.

Trong những năm gần đây, lĩnh vự trí tuệ nhân tạo đã và đang phát triển một cách nhanh chóng, mạnh mẽ và đạt được nhiều thành tựu cao. Điều đó đã tiến gần hơn đến việc trang bị cho máy móc có khả năng của con người như nghe, nhìn và đưa ra quyết định như con người. Cụ thể hơn là việc nhận dạng vật thể, đưa ra dự đoán, ra quyết định và phân loại như con người.

Dựa vào sự tiến bộ này, việc áp dụng lĩnh vực trí tuệ nhân tạo vào nhận dạng các loài hoa đã trở nên phổ biến hơn. Đề tài nghiên cứu “Nhận dạng hoa sử dụng mô hình học sâu” đã được đưa ra nhằm ứng dụng mô hình học sâu vào việc nhận dạng một số loài hoa phổ biến trong cuộc sống. Nghiên cứu sử dụng tập dữ liệu đã được gán nhãn để thực hiện huấn luyện mô hình.

* 1. **Lịch sử giải quyết vấn đề**

Hoa là một trong những sáng tạo đẹp đẽ của thượng đế và chúng tồn tại đến hàng triệu loài và màu sắc khác nhau. Việc xác định từng loài trong số chúng đòi hỏi cần phải có một nhà thực vật học với rất nhiều kiến thức và kỹ năng. Trong thời đại công nghệ đang lên này, hầu hết các vấn đề nan giải đều có thể được thực hiện dễ dàng bằng các kết hợp trí tuệ nhân tạo vào giải quyết các vấn đề trong thế giới thực. Qua đó, các thuật toán học máy như mạng lưới thần kinh tích chập ra đời dùng để xác định các loài hoa chỉ với một hình ảnh sẽ giúp ích rất nhiều cho các ngành như dược phẩm và mỹ phẩm. Ayurveda cổ đại công nhận một số lợi ích chữa bệnh trong hầu hết các loài hoa, điều này càng chứng tỏ mức quan trọng của việc phân biệt từng loài hoa. Từ đó, có nhiều bài báo khoa học ra đời nhằm đưa ra các nghiên cứu về nhiều cách nhận dạng khác nhau.

Nghiên cứu “FLOWER RECOGNITION SYSTEM USING CNN” của tác giả Parvathy, N Vrinda Rao, Shahistha Bai, Naeema Nazer, Prof. Anju [1]. Nội dung của bài báo của nhóm tác giả đã sử dụng mô hình CNN để nhận dạng hoa trên tập dữ liệu “Oxford 102 flowers dataset” mô hình đạt được độ chính xác tổng thể là 90%.

Nghiên Cứu “A FLOWER RECOGNITION SYSTEM USING DEEP NEURAL NETWORK COUPLED WITH VISUAL GEOMETRY GROUP 19 ARCHITECTURE” của các tác giả Ong Zi Yuan, Kah Kien Chye, Huay Wen Kang, Chi Wee Tan[2]. Nhóm tác giả đã sử dụng mô hình VGG19 và ResNet50 để nhận dạng hoa trên tập dữ liệu Flower-102 với độ chính xác lần lượt là VGG19 88% và ResNet50  84%.

Bài Báo “Inception-v3 for flower classification” của nhóm tác giả Xiaoling Xia, Cui Xu, Bing Nan [3] đã sử dụng mô hình Inception-v3 để nhận dạng Hoa trên tập dữ liệu Oxford 102 với độ chính xác 88,33% và trên tập dữ liệu Oxford 17 là 95%.

* 1. **Mục tiêu đề tài**

Nghiên cứu sẽ tập trung vào việc tìm hiểu và ứng dụng mô hình học sâu vào bài toán nhận diện hoa. Sử dụng các mô hình mạng nơ-ron tích chập (CNN) vào việc phân tích, huấn luyện, so sánh độ chính xác giữa các mô hình để đưa ra kết luận. Sau đó cải thiện mô hình tốt nhất để phù hợp với yêu cầu bài toán và xây dựng thành hệ thống với kết quả nhận dạng là tên của hoa.

* 1. **Đối tượng và phạm vi nghiên cứu**

*Đối tượng nghiên cứu:*

* Dữ liệu nghiên cứu là tập các hình ảnh của các loại hoa phổ biến.
* Cấu trúc mạng học sâu trên nhiều lớp tích chập (ConvNet) như: Efficientnet-B5, Resnet-50, Inception-V3.
* Tìm hiểu các thư viện, công cụ hỗ trợ như TensorFlow, Keras,..

*Phạm vi nghiên cứu:*

Đề tài nghiên cứu phương pháp nhận biết một số loài hoa phổ biến hiện nay bằng phương pháp tiếp cận mạng nơ-ron tích chập (CNN). Xây dựng hệ thống nhận diện giúp người dùng sử dụng ảnh chụp đưa vào để phân tích và trả về kết quả là tỉ lệ % về tên của loài hoa đó.

* 1. **Phương pháp nghiên cứu**

Để giải quyết các vấn đề có liên quan đến đề tài Nhận dạng hoa sử dụng mô hình học sâu, các phương pháp nghiên cứu đã được thực hiện như:

*Phương pháp nghiên cứu lý thuyết:*

* Tìm hiểu tài liệu về cách sử dụng, hoạt động của TensorFlow, Keras.
* Tìm hiểu và nghiên cứu kiến trúc mạng học sâu hiện đại, thị giác máy tính.
* Nghiên cứu các mô hình Efficientnet-B5, Resnet-50, Inception-V3 để biết được mô hình nào phù hợp với bài toán.
* Tham khảo các công trình, sản phẩm đã có sẵn và tái hiện lại để hiểu rõ phương pháp, quy trình hoạt động.
* Vận dụng kiến thức đã được học như: kiến thức về lập trình Python, lập trình hướng đối tượng, máy học,…

*Phương pháp thu thập dữ liệu:* Dữ liệu được thu tập từ các nguồn như tải từ các trang web hoặc tự thu thập.

*Phương pháp nghiên cứu thực tiễn:*

* Tiến hành xử lý dữ liệu đã thu thập, xây dựng hệ thống bằng Python.
* Phân chia tập dữ liệu, sử dụng các mô hình CNN để trích xuất đặt trưng, tiến hành phân lớp và huấn luyện mô hình.
* Phân tích, so sánh độ chính xác và đánh giá các phương pháp nhận dạng của các mô hình CNN trên tập dữ liệu thu được.
* Đưa ra kết luận và cải thiện mô hình tốt nhất để phù hợp với tập dữ liệu và yêu cầu của bài toán.
  1. **Kết quả đạt được**

Xây dựng thành công hệ thống “Nhận dạng hoa sử dụng mô hình học sâu” với 25 loại hoa phổ biến hiện nay với chức năng sử dụng ảnh có sẵn để nhận dạng và hiển thị tỉ lệ phần trăm (%) độ chính xác tên của từng loài hoa.

* 1. **Bố cục luận văn**

Nội dung của quyển luận văn bao gồm 05 chương:

* Chương 1. Giới thiệu: Giới thiệu tổng quát về đề tài “Nhận dạng hoa sử dụng mô hình học sâu”, nêu lịch sử giải quyết đề tài, mục tiêu chính của đề tài, đối tượng và phạm vi nghiên cứu, phương pháp nghiên cứu đề tài và kết quả đạt được.
* Chương 2. Cơ sở lý thuyết: Tìm hiểu về thị giác máy tính, mô hình học sâu, mô hình mạng nơ-ron tích chập, các phương pháp tiếp cận trong nhận dạng hoa, cã thư viện hỗ trợ.
* Chương 3. Phương pháp thực hiện: Trình bày tổng quan các bước và cách cài đặt giải thuật.
* Chương 4. Kết quả thực nghiệm: Trình bày kết quả kiểm tra và đánh giá kết quả đạt được.
* Chương 5. Kết luận và hướng phát triển: Nêu kết quả đạt được và hướng phát triển của đề tài.

# **CHƯƠNG 2. CƠ SỞ LÝ THUYẾT**

Chương 2 sẽ trình bày các cơ sở lý thuyết có liên quan đến đề tài “Nhận dạng hoa sử dụng mô hình học sâu”, gồm:

* Thị giác máy tính (Computer Vision)
* Mô hình học sâu (Deep Learning)
* Mô hình mạng nơ-ron tích chập (Convolutional Neural Network)
* Các phương pháp tiếp cận trong nhận dạng hoa
  1. **Thị giác máy tính**

Theo Wikipedia, Thị giác máy tính (Computer Vision) là một lĩnh vực bao gồm các phương pháp thu nhận, [xử lý ảnh kỹ thuật số](https://vi.wikipedia.org/w/index.php?title=X%E1%BB%AD_l%C3%BD_%E1%BA%A3nh_k%E1%BB%B9_thu%E1%BA%ADt_s%E1%BB%91&action=edit&redlink=1), phân tích và nhận dạng các hình ảnh và, nói chung là dữ liệu đa chiều từ thế giới thực để cho ra các thông tin số hoặc biểu tượng, ví dụ trong các dạng quyết định. Việc phát triển lĩnh vực này có bối cảnh từ việc sao chép các khả năng thị giác con người bởi sự nhận diện và hiểu biết một hình ảnh mang tính điện tử. Sự nhận diện hình ảnh có thể xem là việc giải quyết vấn đề của các biểu tượng thông tin từ dữ liệu hình ảnh qua cách dùng các mô hình được xây dựng với sự giúp đỡ của các ngành [lý thuyết học](https://vi.wikipedia.org/w/index.php?title=L%C3%BD_thuy%E1%BA%BFt_h%E1%BB%8Dc&action=edit&redlink=1), [thống kê](https://vi.wikipedia.org/wiki/Th%E1%BB%91ng_k%C3%AA), [vật lý](https://vi.wikipedia.org/wiki/V%E1%BA%ADt_l%C3%BD) và [hình học](https://vi.wikipedia.org/wiki/H%C3%ACnh_h%E1%BB%8Dc). Thị giác máy tính cũng được mô tả là sự tổng thể của một dải rộng các quá trình tự động và tích hợp và các thể hiện cho các nhận thức thị giác.

Thị giác máy tính là một lĩnh vực mới mẽ và đang phát triển một các vượt bật trên toàn thế giới. Đây là một lĩnh vực tập trung tái tạo lại hệ thống thị giác của con người, từ đó cho phép máy tính có thể xử lý nhằm xác định, nhận diện ra các đối tượng trong hình ảnh cũng như video như con người. Đặt biệt, lượng dữ liệu được sử dụng trong việc đào tạo khá lớn, điều này góp phần giúp cho tỷ lệ chính xác của nhận dạng đối tượng ngày càng tăng. Điều này cũng đã góp phần làm cho thị giác máy tính ngày càng phát triển.

Thị giác máy tính có nhiều lĩnh vực như: Xử lý hình ảnh, Nhận diện mẫu, Quang trắc. Trong đó, xử lý hình ảnh là một trong những lĩnh vực quan trọng trong thị giác máy tính. Nó tuy là một ngành khoa học khá mới mẻ nhưng tốc độ phát triển lại vô cùng mạnh mẽ và mở ra nhiều tiềm năng cho các lĩnh vực nghiên cứu cho Việt Nam nói chung và thế giới nói riêng.

* 1. **Mô hình học sâu**

Theo Wikipedia, Học sâu (Deep Learning) còn gọi là học cấu trúc sâu, là một phần trong một nhánh rộng hơn các phương pháp học máy dựa trên mạng thần kinh nhân tạo kết hợp với việc học biểu diễn đặc trưng (Representation Learning). Việc học này có thể có giám sát, nửa giám sát hoặc không giám sát. Học sâu thường được nhắc đến cùng với dữ liệu lớn (Big Data) và trí tuệ nhân tạo (AI). Đã có nhiều ứng dụng trong thực tế, đang phát triển mạnh mẽ theo sự phát triển của tốc độ máy tính đặt biệt là khả năng tính toán trên GPU và sự tăng nahnh của dữ liệu cùng với framework (TensorFow hay Pytorch) làm việc xây dựng model trở nên dễ dàng hơn.

Các mô hình học sâu gồm: Mạng nơ-ron tái tạo (Recurrent Neural Networks), mạng nơ-ron nhân tạo (Autoencoder), mạng nơ-ron sâu (Deep Neural Network), mạng nơ-ron tích chập (Convolutional Neural Network) và mạng học sâu niềm tin (Deep belief Net).

* 1. **Mô hình mạng nơ-ron tích chập**

Mạng CNN ra đời đã thúc đẩy quá trình phát triển của ngành computer vision. Hiện tại có rất nhiều các kiến trúc mạng CNN khác nhau và các kiến trúc mới vẫn đang tiếp tục được khám phá ngày qua ngày. Nhưng ít ai biết rằng đằng sau những khám phá đó là một tiến trình khoa học lâu dài và bền bỉ trong gần 20 năm. Với sự kết hợp đồng bộ của phát triển kiến trúc mạng, khả năng tính toán của máy tính và các phương pháp tối ưu hóa.

Theo Wikipedia, mạng thần kinh tích chập (còn gọi là mạng nơ-ron tích chập hay ít phổ biến hơn là mạng thần kinh/nơ-ron chuyển đổi, [tiếng Anh](https://vi.wikipedia.org/wiki/Ti%E1%BA%BFng_Anh): convolutional neural network, viết tắt CNN hay ConvNet) là một lớp của [mạng thần kinh sâu](https://vi.wikipedia.org/w/index.php?title=M%E1%BA%A1ng_th%E1%BA%A7n_kinh_s%C3%A2u&action=edit&redlink=1) (deep neural network), áp dụng phổ biến nhất để phân tích hình ảnh trực quan. Mạng còn được gọi là shift invariant (dịch chuyển bất biến) hay mạng thần kinh nhân tạo không gian bất biến (SIANN), dựa trên kiến trúc trọng số được chia sẻ và các đặc tính [đối xứng tịnh tiến](https://vi.wikipedia.org/w/index.php?title=%C4%90%E1%BB%91i_x%E1%BB%A9ng_t%E1%BB%8Bnh_ti%E1%BA%BFn&action=edit&redlink=1) (translational symmetry). CNN có nhiều ứng dụng trong [thị giác máy tính](https://vi.wikipedia.org/wiki/Th%E1%BB%8B_gi%C3%A1c_m%C3%A1y_t%C3%ADnh), [hệ thống gợi ý](https://vi.wikipedia.org/wiki/H%E1%BB%87_th%E1%BB%91ng_g%E1%BB%A3i_%C3%BD), [phân loại hình ảnh](https://vi.wikipedia.org/w/index.php?title=Ph%C3%A2n_lo%E1%BA%A1i_h%C3%ACnh_%E1%BA%A3nh&action=edit&redlink=1), [tính toán hình ảnh y tế](https://vi.wikipedia.org/w/index.php?title=T%C3%ADnh_to%C3%A1n_h%C3%ACnh_%E1%BA%A3nh_y_t%E1%BA%BF&action=edit&redlink=1) ([điện toán](https://vi.wikipedia.org/wiki/%C4%90i%E1%BB%87n_to%C3%A1n) hình ảnh y tế), [xử lý ngôn ngữ tự nhiên](https://vi.wikipedia.org/wiki/X%E1%BB%AD_l%C3%BD_ng%C3%B4n_ng%E1%BB%AF_t%E1%BB%B1_nhi%C3%AAn), và [chuỗi thời gian](https://vi.wikipedia.org/wiki/Chu%E1%BB%97i_th%E1%BB%9Di_gian) tài chính.

Mạng nơ-ron tích chập được ứng dụng rất nhiều trong học sâu, trong thị giác máy tính và mô hình hóa âm thanh cho nhận dạng giọng nói tự động. Cấu trúc một mạng CNN đơn giản (Hình 2.1) gồm:

Input -> Convolution + Relu -> Pooling -> Fully-Connected -> Output

Trong đó:

Input: Lớp đầu vào

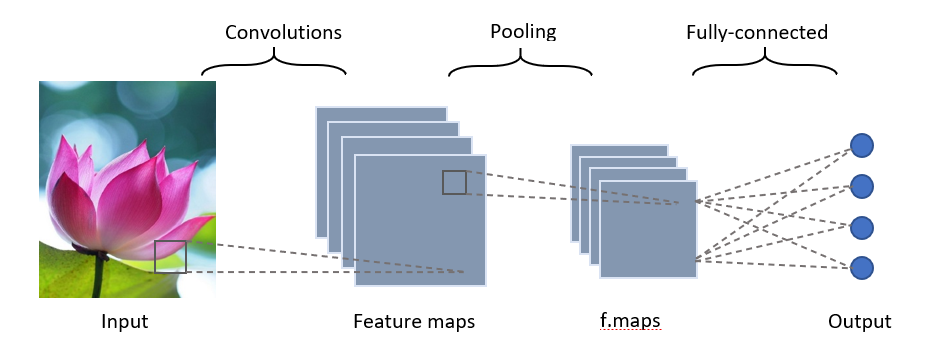
Convolution: Lớp tích chập

Relu: Hàm kích hoạt phi tuyến

Pooling: Lớp tổng hợp

Fully-connected: Lớp kết nối hoàn toàn

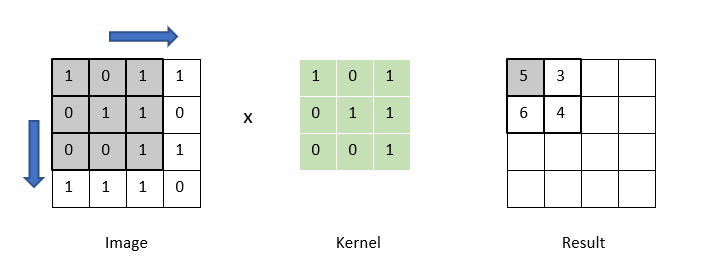
Output: Đơn vị đầu ra đại diện



Hình 2.1. Cấu trúc mạng nơ-ron tích chập đơn giản

**Lớp tích chập (Convolution Layer)**

Các lớp tích chập có các kernel đã được học để tự lấy ra các thông tin chính xác nhất mà không cần chọn các feature. Lớp tích chập sử dụng bộ lọc để thực hiện phép tích chập khi đi qua các ma trận lân cận cùng kích thước với chiều từ trái sang phải, từ trên xuống dưới. Kết quả đầu ra được gọi là feature hay activation map. Số lớp tích chập càng nhiều thì hiệu suất mô hình càng được cải thiện.



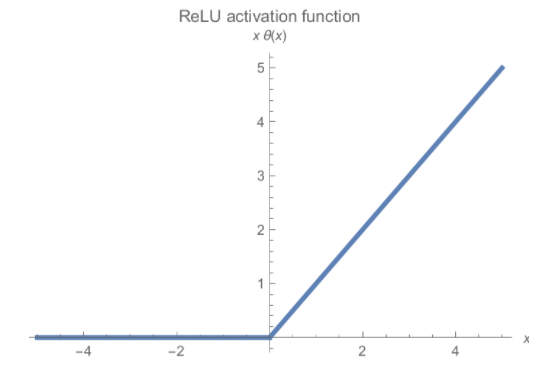
Hình 2.2. Minh họa phép tích chập

Trong minh họa phép tích chập Hình 2.2*,* đầu vào là một ma trận có kích thước 4x4 tương ứng với hình ảnh, một bộ lọc (kernel) có kích thước 3x3 dùng để biến đổi ma trận đầu vào. Bộ lọc sẽ dịch chuyển trên ma trận đầu vào và thực hiện phép tích chập. Phép tích chập đơn giản là phép nhân và phép cộng giữa ma trận đầu vào với bộ lọc. Kết quả sau khi tích chập là một ma trận có kích thước 2x2.

**Hàm kích hoạt phi tuyến ReLU**

Sau mỗi lớp tích chập, đầu ra của mỗi lớp có thể có các giá trị âm không cần thiết, các giá trị này có thể làm ảnh hưởng cho việc tính toán ở các layer sau. Do đó, các hàm phi tuyến được áp dụng lên đầu ra của mỗi lớp để đưa các giá trị âm về thành 0. Một số hàm tuyến tính được sử dụng như ReLU, tanh, sigmoid, … Trong đó, hàm kích hoạt ReLU cho hiệu suất tốt nhất, việc sử dụng hàm kích hoạt Relu không làm thay đổi kích thước của ảnh và cũng không cần thêm tham số nào khác. Hàm ReLu được định nghĩa như Hình 2.3:

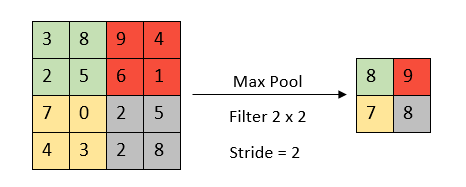
f(x) = max (0, x)



Hình 2.3. Hàm kích hoạt phi tuyến ReLU

**Lớp kết hợp (Pooling Layer)**

Lớp kết hợp thường được sử dụng giữa các lớp tích chập để giảm chiều của không gian đầu vào và giảm bớt số lượng tham số khi hình ảnh quá lớn nhưng vẫn giữ được các thuộc tính quan trọng. Lớp kết hợp còn giúp kiểm soát hiện tượng overffiting. Các lớp kết hợp thường sử dụng kích thước 2x2 hoặc 4x4 cho ảnh đầu vào lớn. Thông thường, có nhiều hình thức Pooling khác nhau cho nhiều bài toán, nhưng phép Pooling được sử dụng phổ biến là Max Pooling dùng để chọn ra giá trị lớn nhất trong vùng đầu vào, kích thước filter thường dùng là 2x2 với giá trị Stride = 2 (được minh họa ở Hình 2.4), ngoài ra còn có Average Pooling và Global Pooling.



Hình 2.4. Minh họa phép Max Pooling trên ma trận 4x4

với Filter = 2x2 và Stride = 2, kết quả đầu ra là ma trận 2x2

**Lớp kết nối hoàn toàn (Fully-connected Layer)**

Sau khi ảnh được truyền qua nhiều lớp tích chập và lớp kết hợp, bức ảnh sẽ được dàn phẳng thành 1 vector thay vì mảng nhiều chiều như trước và đưa vào một lớp được kết nối như một mạng nơ-ron. Lớp kết nối hoàn toàn có nhiệm vụ kết nối các nơ-ron tại lớp này với tất cả nơ-ron của lớp trước đó tạo thành một mạng nơ-ron với đầu ra gồm các nơ-ron có số lượng tùy vào bài toán.

* 1. **Các phương pháp tiếp cận trong nhận dạng hoa**

1. 4. 1. **Mô hình Efficientnet-B5**

Trong bài báo ICML 2019 "EfficientNet: Rethinking Model Scaling for Convolutional Neural Net- works" nhóm tác giả Mingxing Tan, Quoc V. Le đã đề xuất một phương pháp chia tỷ lệ mới giúp chia tỉ lệ đồng đều tất cả các kích thước chiều sâu/ chiều rộng/độ phan giải bằng cách sử dụng hệ số ghép đơn giản nhưng hiệu quả cao để mở rộng quy mô được gọi là Efficientnet - là một mạng cơ sở kiến trúc di động.[4]

Mô hình Efficientnet-B5 là mộ trong những mô hình Efficientnet được tác giả thiết kể để phân loại hình ảnh và Efficientnet-B5 đã được đào tạo trong TensorFlow và trên cơ sở dữ liệu hình ảnh ImageNet

* + 1. **Mô hình Resnet-50**

Mạng ResNet là viết tắt của Residual Network, đây là mạng nơ-ron được nhóm nghiên cứu của Microsoft Research bao gồm các thành viên: Kaiming He, Xiangyu Zhang, Shaoqing Ren và Jian Sun sáng kiến giới thiệu vào năm 2015 trong một nghiên cứu về thị giác máy tính với bài báo được mang tên “Deep Residual Learning for Image Recognition”. Với nghiên cứu ở mô hình mạng ResNet nhóm nghiên cứu đã giành được vị trí cao nhất tại cuộc thi phân loại ILSVRC 2015 với tỉ lệ lỗi chỉ 3,57%. Ngoài ra, nó cũng đứng đầu trong việc phát hiện hay định vị ImageNet, phát hiện COCO và phân đoạn COCO trong các cuộc thi ILSVRC & COCO năm 2015.

Diagram

Description automatically generated

Hình 2.5. Kiến trúc ResNet bao gồm 2 khối đặt trưng

là khối tích chập (Conv Block) và khối xác định (Identity Block)

ResNet là kiến trúc được sử dụng phổ biến nhất ở thời điểm hiện tại. ResNet cũng là kiến trúc sớm nhất áp dụng batch normalization. Mặc dù là một mạng rất sâu khi có số lượng layer lên tới 152 nhưng nhờ áp dụng những kỹ thuật đặc biệt mà ta sẽ tìm hiểu bên dưới nên kích thước của ResNet50 chỉ khoảng 26 triệu tham số. Kiến trúc với ít tham số nhưng hiệu quả của ResNet đã mang lại chiến thắng trong cuộc thi ImageNet năm 2015.

Những kiến trúc trước đây thường cải tiến độ chính xác nhờ gia tăng chiều sâu của mạng CNN. Nhưng thực nghiệm cho thấy đến một ngưỡng độ sâu nào đó thì độ chính xác của mô hình sẽ bão hòa và thậm chí phản tác dụng và làm cho mô hình kém chính xác hơn. Khi đi qua quá nhiều tầng độ sâu có thể làm thông tin gốc bị mất đi thì các nhà nghiên cứu của Microsoft đã giải quyết vấn đề này trên ResNet bằng cách sử dụng kết nối tắt.

Các kết nối tắt (skip connection) giúp giữ thông tin không bị mất bằng cách kết nối từ layer sớm trước đó tới layer phía sau và bỏ qua một vài layers trung gian. Trong các kiến trúc base network CNN của các mạng YOLOv2, YOLOv3 và gần đây là YOLOv4 bạn sẽ thường xuyên thấy các kết nối tắt được áp dụng.

ResNet có khối tích chập (Convolutional Bock, chính là Conv block trong hình) sử dụng bộ lọc kích thước 3 x 3 giống với của InceptionNet. Khối tích chập bao gồm 2 nhánh tích chập trong đó một nhánh áp dụng tích chập 1 x 1 trước khi cộng trực tiếp vào nhánh còn lại.

Khối xác định (Identity block) thì không áp dụng tích chập 1 x 1 mà cộng trực tiêp giá trị của nhánh đó vào nhánh còn lại.

Diagram

Description automatically generated

Hình 2.6. Cộng trực tiếp đầu vào của khối với nhánh còn lại trong khối Identity block

Giả sử chúng ta có x là đầu vào của khối xác định. Chúng ta cần ánh xạ đầu vào x thành hàm f(x). Để tìm ra ánh xạ chuẩn xác tương đương với hàm f(x) là một việc khá khó. Nhưng nếu cộng thêm ở đầu ra thành x+f(x) thì chúng ta sẽ qui về tham số hóa độ lệch, tức cần tham số hóa phần dư f(x). Tìm ánh xạ theo phần dư sẽ dễ hơn nhiều vì chỉ cần tìm giá trị f(x) sao cho nó gần bằng 0 là có thể thu được một ánh xạ chuẩn xác. Tại một khối xác định, chúng ta sẽ áp dụng một layer activation ReLU sau mỗi xen kẽ giữa những tầng trọng số.

Mặc dù có kiến trúc khối kế thừa lại từ GoogleNet nhưng ResNet lại dễ tóm tắt và triển khai hơn rất nhiều vì kiến trúc cơ sở của nó chỉ gồm các khối tích chập và khối xác định. Ta có thể đơn giản hóa kiến trúc của ResNet-50 như Hình 2.7 bên dưới:

Diagram

Description automatically generated

Hình 2.7. Kiến trúc tóm tắt của mạng Resnet-50

* + 1. **Mô hình Inception-V3**

Mạng Inception là một mạng nơ-ron hỗ trợ phân tích hình ảnh và phát hiện đối tượng, được các nhà nghiên cứu của Google đưa ra vào năm 2014 để tham gia cuộc thi ImageNet. Mạng Inception chứa các đơn vị gọi là inception cell thực hiện tích chập 1 đầu vào với nhiều bộ lọc khác nhau và tổng hợp lại theo nhiều nhánh, vì vậy các tầng nối tiếp không hoàn toàn gối đầu lên nhau. Với mỗi cell, các bộ lọc có nhiều kích thước khác nhau để trích các đặc trưng từ đầu vào.

Hình 2.8 bên dưới cho thấy kiến trúc của mô hình Inception-V3:

Diagram, schematic

Description automatically generated

Hình 2.8. Kiến trúc GoogleNet – Inception version 3

Inception-V3 là kế thừa của Inception-V1 bao gồm 24 triệu tham số. Toàn bộ các layer tích chập của Inception-V3 được theo sau bởi một layer batch normalization và một ReLU activation. Batch normalization là kỹ thuật chuẩn hóa đầu vào theo từng minibatch tại mỗi layer theo phân phối chuẩn hóa N(0,1), giúp cho quá trình huấn luyện thuật toán nhanh hơn.

Inception-V3 giải quyết được vấn đề thắt cổ chai (representational bottlenecks). Tức là kích thước của các layers không bị giảm một cách đột ngột. Đồng thời Inception-V3 có một cách tính toán hiệu quả hơn nhờ sử dụng phương pháp nhân tố (factorisation methods).

Hiện tại Inception module bao gồm 4 version. Chúng ta hãy cùng xem qua các điểm đặc biệt ở từng version:

* Inception-V1: Khối Inception sẽ bao gồm 4 nhánh song song. Các bộ lọc kích thước lần lượt là 1x1, 3x3, 5x5 được áp dụng trong Inception Module giúp trích lọc được đa dạng đặc trưng trên những vùng nhận thức có kích thước khác nhau. Ở đầu các nhánh 1, 2, 4 từ trên xuống, phép tích chập 1x1 được sử dụng trên từng điểm ảnh như một kết nối fully connected nhằm mục đích giảm độ sâu kênh và số lượng tham số của mô hình. Nhánh thứ 3 từ trên xuống chúng ta giảm chiều dữ liệu bằng một layer max-pooling kích thước 3x3 và sau đó áp dụng bộ lọc kích thước 1x1 để thay đổi số kênh.
* Inception-A: Cải tiến so với Inception module V1. Tại nhãnh thứ nhất thay 1 layer tích chập 5 x 5 bằng 2 layer tích chập 3 x 3 liên tiếp giúp giảm số lượng tham số từ 25 về 18 và tăng độ sâu cho mô hình.
* Inception-B: Cải tiến so với Inception-A. Thay tích chập 3 x 3 bằng tích chập 7 x 7 ở nhánh thứ nhất và nhánh thứ 2. Đồng thời chúng ta phân tích nhân tố tích chập 7 x 7 thành 2 tích chập liên tiếp 7 x 1 và 1 x 7 số lượng tham số sẽ ít hơn so với tích chập 2 tích chập 3 x 3 liên tiếp. Nhờ đó số lượng tham số giảm từ 18 xuống còn 14.
* Inception-C: Cải tiến so với Inception-B. Thay tích chập 7 x 1 bằng tích chập 3 x 1 và 1 x 7 bằng 1 x 3 và đồng thời thay vì đặt layer 3 x 1 và 1 x 3 liên tiếp thì đặt chúng song song. Kiến trúc này giúp giảm số lượng tham số từ 14 về còn 6.

Ngoài ra ở Inception-V3 chúng ta còn sử dụng 2 kiến trúc giảm chiều dữ liệu là Reduction-A và Reduction-B.

# **CHƯƠNG 3. PHƯƠNG PHÁP THỰC HIỆN**

Chương 3 sẽ trình bày tổng quan về các bước thực hiện và cách cài đặt giải thuật. Các mô hình được sử dụng là Inception-V3, Resnet-50 và Efficientnet-B5.

* 1. **Tổng quan**

Diagram

Description automatically generated

Hình 3.1. Sơ đồ các bước thực hiện

Hình 3.1 là sơ đồ các bước thực hiện công việc, cụ thể như sau:

Bước 1. Thu thập dữ liệu: Tiến hành thu thập dữ liệu hình ảnh về các loài hoa cần nhận diện. Hình ảnh thu thập tốt nhất là những hình ảnh rõ ràng không bị mờ.

Bước 2. Xử lý dữ liệu: Xử lý hình ảnh vừa thu thập được bao gồm việc loại bỏ những hình ảnh mờ không đạt tiêu chuẩn, sửa tên hình ảnh, gán nhãn và phân chia tập dữ liệu thành tập train và validation.

Bước 3. Huấn luyện các mô hình: Sử dụng Google Colaboratory để chạy huấn luyện lần lượt các mô hình như 2 mô hình CNN tự xây dựng, mô hình Inception-V3, Resnet-50 và Efficientnet-B5.

Bước 4. Đánh giá và chọn mô hình tốt nhất cho bài toán: Sau khi huấn luyện mô hình, tiến hành so sánh hiệu quả dựa vào độ chính xác của từng mô hình. Mô hình nào có độ chính xác cao sẽ được lấy để tiến hành bước tiếp theo.

Bước 5. Xử lý lại mô hình được chọn: Sau khi chọn được mô hình hiệu quả, tiến hành xử lý lại mô hình để phù hợp với tập dữ liệu và yêu cầu bài toán.

Bước 6. Tải mô hình về máy và thiết kế giao diện: Bước cuối cùng là tải mô hình đã qua xử lý về máy và tiến hành thiết kế giao diện. Xây dựng ứng dụng web sử dụng Python Flask, ứng dụng sử dụng ảnh trực tiếp và nhận diện tên của hình ảnh hoa được upload.

* 1. **Cài đặt giải thuật**
     1. **Thu thập dữ liệu**

Tập dữ liệu hoa bao gồm 1433 hình ảnh của 25 loài hoa khác nhau, là những loài hoa thân thuộc, phổ biến, thường được trồng quanh nhà, ven đường hay các khu du lịch,… Tập dữ liệu được thu thập từ nhiều nguồn khác nhau như được tải về từ các tragn web, tự chụp,… Các ảnh trong tập dữ liệu đều đã được loại bỏ những ảnh mờ, không đúng định dạng và được chuyển về dạng “.jpg”.

Hình 3.2 và Hình 3.3 là hình ảnh về một số loài hoa cụ thể như sau:



Hình 3.2. Hình ảnh ví dụ của Hoa Cẩm Chướng

A plant in a pot

Description automatically generated with medium confidence

Hình 3.3. Hình ảnh ví dụ của Hoa Đồng Tiền

Bảng 3.1 Thống kê số lượng các loài hoa trong tập dữ liệu, cụ thể như sau:

Bảng 3.1. Tập dữ liệu hình ảnh

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Tên hoa** | **Số lượng ảnh** | **STT** | **Tên hoa** | **Số lượng ảnh** |
| 1 | Hoa Bướm Bạc | 58 | 14 | Hoa Lan | 58 |
| 2 | Hoa Cẩm Chướng | 60 | 15 | Hoa Layon | 59 |
| 3 | Hoa Chi Dây Huỳnh | 57 | 16 | Hoa Ly | 58 |
| 4 | Hoa Cúc | 57 | 17 | Hoa Mai | 59 |
| 5 | Hoa Dâm Bụt | 58 | 18 | Hoa Mười Giờ | 47 |
| 6 | Hoa Đào | 59 | 19 | Hoa Nhài | 63 |
| 7 | Hoa Dã Quỳ | 58 | 20 | Hoa Sen | 59 |
| 8 | Hoa Đồng Tiền | 57 | 21 | Hoa Thược Dược | 58 |
| 9 | Hoa Giấy | 57 | 22 | Hoa Tigon | 67 |
| 10 | Hoa Hồng | 59 | 23 | Hoa Trang | 45 |
| 11 | Hoa Hồng Môn | 57 | 24 | Hoa Tulip | 57 |
| 12 | Hoa Huệ | 58 | 25 | Hoa Tường Vi | 53 |
| 13 | Hoa Hướng Dương | 55 |  |  |  |

Chia tập dữ liệu thành 2 phần dùng để huấn luyện các mô hình:

* Dùng 1147 ảnh dùng để huấn luyện mô hình.
* Dùng 286 ảnh còn lại dùng để kiểm tra với tỷ lệ tập kiểm tra là 0.2 (20%), đồng thời cũng dùng để kiểm nghiệm và đánh giá độ chính xác của các mô hình sau khi huấn luyện.
  + 1. **Huấn luyện mô hình**

Do máy không có hỗ trợ GPU và lượng dữ liệu cũng khá lớn nên các mô hình sẽ được huấn luyện trên Google Colaboratory. Sau khi chuẩn bị xong dữ liệu và mô hình thì tiếp hành tải lên Google Drive để tiến hành huấn luyện.

Các bước thiết lập Google Colab được thực hiện như sau:

* Gắn kết ổ đĩa
* Kiểm tra việc gắn kết ổ đĩa
* Download và cài đặt các gói thư viện cần thiết
* Sử dụng GPU để huấn luyện mô hình
* Huấn luyện các mô hình CNN, Inception-V3, Resnet-50, Efficientnet-B5
* Xử lý lại mô hình, ở đây mô hình được chọn là Inception-V3
* Xuất mô hình Inception-V3

Sau khi huấn luyện các mô hình CNN, Inception-V3, Resnet-50, Efficientnet-B5 để tiến hành so sánh thì ta có kết quả ban đầu về độ chính xác của từng mô hình qua 10 lần duyệt (epochs = 10) như Bảng 3.2 dưới đây:

Bảng 3.2. Độ chính xác của từng mô hình

|  |  |  |
| --- | --- | --- |
|  | **Accuracy** | **Val\_accuracy** |
| CNN (model\_1) | 0.775 | 0.308 |
| CNN (model\_2) | 0.153 | 0.161 |
| Inception-V3 | 0.951 | 0.629 |
| Resnet-50 | 0.129 | 0.15 |
| Efficientnet-B5 | 0.059 | 0.059 |

Mô hình Inception-V3 hoạt động tốt hơn các mô hình còn lại. Do đó, mô hình Inception-V3 sẽ được chọn để xây dựng ứng dụng nhận dạng hoa.

* + 1. **Thiết kế giao diện**

Sau khi chọn xong mô hình, bước tiếp theo là xử lý lại mô hình để phù hợp hơn với yêu cầu bài toán, sau đó xuất mô hình về về máy và tiến hành thiết kế giao diện. Sử dụng Python, Flask để tạo một ứng dụng web đơn giản và sử dụng javascript để xử lý. Fask là một microframework Python để tạo ứng dụng web.

Hình 3.4 cho thấy giao diện ban đầu của ứng dụng web khi load trang:

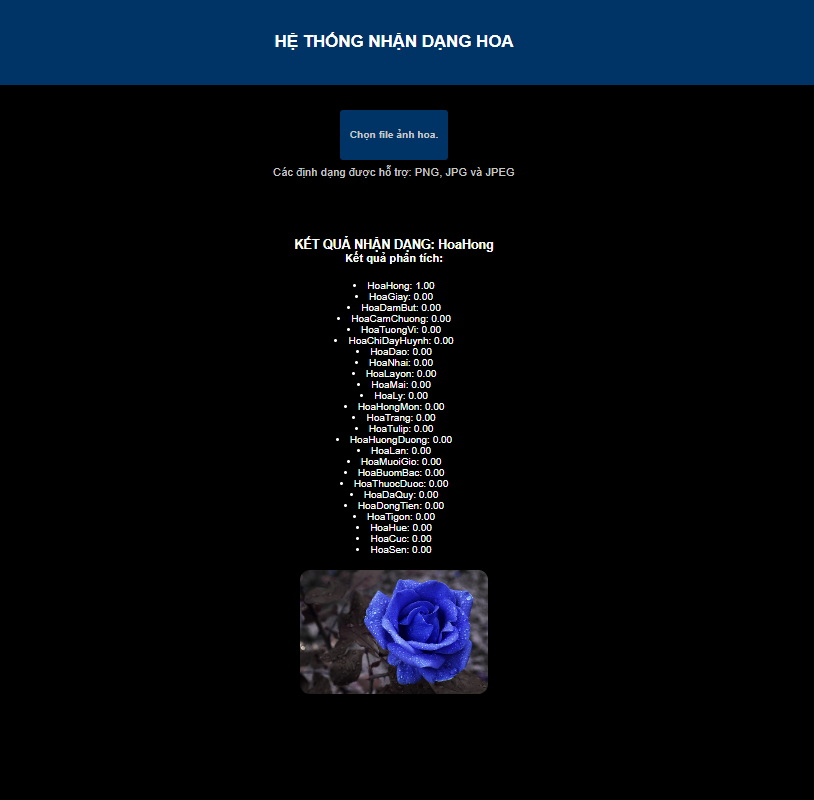
Graphical user interface, application

Description automatically generated

Hình 3.4. Giao diện của ứng dụng web khi load trang

Chức năng tổng quát của mô hình là nhận dạng đối tượng hoa thông qua dữ liệu mô hình được huấn luyện trước, cái bước thực hiện như sau:

* Bước 1: Tại giao diện của website, người dùng tiến hành chọn ảnh cần nhận dạng và tiến hành tải ảnh lên, dữ liệu sẽ được lưu lại và gửi thông tin về máy chủ để xử lý
* Bước 2: Server khi nhận được dữ liệu sẽ tiến hành nhận dạng đối tượng, kết quả trả về là một ảnh và phần trăm (%) độ chính xác của ảnh đó.
* Bước 3: Server trả kết quả về giao diện web, gồm ảnh và % độ chính xác (Hình 3.5)



Hình 3.5. Giao diện kết quả trả về sau khi upload hình ảnh lên web

# **CHƯƠNG 4. KẾT QUẢ THỰC NGHIỆM**

1. **Kết quả kiểm tra**

Chỉ số độ chính xác tiêu chuẩn cho các hệ thống phát hiện đối tượng là độ chính xác trung bình (mean Average Precision – mAP). Trong phần này, được chia thành các phần tối thiểu và được mô tả như sau:

* Độ chính xác (Precision):

Precision =

Trong đó:

* True positive (TP): Số lượng các điểm phân loại đúng lớp và được dự đoán là đúng (Mô hình nhận dạng đúng hình ảnh)
* False positive (FP): Số lượng các điểm phân loại sai lớp và được dự đoán là đúng (Mô hình nhận dạng sai hình ảnh)
* Hàm gọi lại (Recall):

Recall =

Với False Negative (FN): Số lượng các điểm phân loại đúng lớp và được dự đoán sai

* Độ do F1-score: Điểm F còn được gọi là điểm F1, là thước đo độ chính xác của mô hình trên tập dữ liệu

F1 = 2 \*

Tiến hành kiểm thử trên từng ảnh sử dụng mô hình Inception-V3 với tập dữ liệu có 25 lớp gồm 250 hình các loài hoa.

Lớp dữ liệu hoa Dâm Bụt (HoaDamBut):

* Tổng số lượng hình ảnh kiểm thử: 20
* Số lượng hình ảnh hoa Dâm Bụt: 10
* Số lượng hình ảnh nhận diện được là hoa Dâm Bụt: 11
* Số lượng hình ảnh nhận diện đúng nhãn hoa Dâm Bụt: 8

Bảng 4.1. Bảng Confusion maxtix cho lớp Hoa Dâm Bụt

|  |  |  |
| --- | --- | --- |
| Dự đoán => | Positive | Negative |
| Positive | True Positive (TP) = 8 | False Positive (FN) =2 |
| Negative | False Negative (FP)= 3 | True Negative (TN) = 7 |

Precision = = = 0.73

Recall = = = 0.8

F1 = 2 \* = 2 \* = 0.76

Bảng 4.2 dưới đây là bảng thống kê số liệu sau khi tiến hành kiểm thử cho từng lớp hoa:

Bảng 4.2. Thống kê số lượng hình ảnh sau kiểm thử

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Số lượng ảnh kiểm thử** | **Số lượng ảnh của nhãn** | **Nhận diện được** | **Nhận diện đúng nhãn** |  | **Số lượng ảnh kiểm thử** | **Số lượng hình** | **Nhận diện được** | **Nhận diện đúng nhãn** |
| Hoa Bướm Bạc | 20 | 10 | 9 | 8 | Hoa Lan | 20 | 10 | 8 | 7 |
| Hoa Cẩm Chướng | 20 | 10 | 11 | 8 | Hoa Layon | 20 | 10 | 10 | 9 |
| Hoa Chi Dây Huỳnh | 20 | 10 | 9 | 8 | Hoa Ly | 20 | 10 | 9 | 8 |
| Hoa Cúc | 20 | 10 | 10 | 8 | Hoa Mai | 20 | 10 | 9 | 8 |
| Hoa Dâm Bụt | 20 | 10 | 11 | 8 | Hoa Mười Giờ | 20 | 10 | 8 | 7 |
| Hoa Đào | 20 | 10 | 10 | 8 | Hoa Nhài | 20 | 10 | 10 | 9 |
| Hoa Dã Quỳ | 20 | 10 | 8 | 7 | Hoa Sen | 20 | 10 | 10 | 9 |
| Hoa Đồng Tiền | 20 | 10 | 10 | 9 | Hoa Thược Dược | 20 | 10 | 10 | 9 |
| Hoa Giấy | 20 | 10 | 13 | 9 | Hoa Tigon | 20 | 10 | 9 | 7 |
| Hoa Hồng | 20 | 10 | 9 | 8 | Hoa Trang | 20 | 10 | 8 | 7 |
| Hoa Hồng Môn | 20 | 10 | 10 | 9 | Hoa Tulip | 20 | 10 | 10 | 8 |
| Hoa Huệ | 20 | 10 | 10 | 9 | Hoa Tường Vi | 20 | 10 | 10 | 9 |
| Hoa Hướng Dương | 20 | 10 | 8 | 7 |  |  |  |  |  |

mAP của 25 lớp sử dụng mô hình Inception-V3 được thống kê như Bảng 4.3:

Bảng 4.3. mAP mô hình Inception-V3

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1-score** |  | **Precision** | **Recall** | **F1** |
| Hoa Bướm Bạc | 0.89 | 0.8 | 0.84 | Hoa Lan | 0.88 | 0.7 | 0.78 |
| Hoa Cẩm Chướng | 0.73 | 0.8 | 0.76 | Hoa Layon | 0.9 | 0.9 | 0.9 |
| Hoa Chi Dây Huỳnh | 0.89 | 0.8 | 0.84 | Hoa Ly | 0.89 | 0.8 | 0.84 |
| Hoa Cúc | 0.8 | 0.8 | 0.8 | Hoa Mai | 0.89 | 0.8 | 0.84 |
| Hoa Dâm Bụt | 0.73 | 0.8 | 0.76 | Hoa Mười Giờ | 0.88 | 0.7 | 0.78 |
| Hoa Đào | 0.8 | 0.8 | 0.8 | Hoa Nhài | 0.9 | 0.9 | 0.9 |
| Hoa Dã Quỳ | 0.88 | 0.7 | 0.78 | Hoa Sen | 0.9 | 0.9 | 0.9 |
| Hoa Đồng Tiền | 0.9 | 0.9 | 0.9 | Hoa Thược Dược | 0.9 | 0.9 | 0.9 |
| Hoa Giấy | 0.69 | 0.9 | 0.78 | Hoa Tigon | 0.78 | 0.7 | 0.74 |
| Hoa Hồng | 0.89 | 0.8 | 0.84 | Hoa Trang | 0.88 | 0.7 | 0.78 |
| Hoa Hồng Môn | 0.9 | 0.9 | 0.9 | Hoa Tulip | 0.8 | 0.8 | 0.8 |
| Hoa Huệ | 0.9 | 0.9 | 0.9 | Hoa Tường Vi | 0.9 | 0.9 | 0.9 |
| Hoa Hướng Dương | 0.88 | 0.7 | 0.78 |  |  |  |  |

Giá trị trung bình mAP (mean Average Precision) của 25 lớp của mô hình Inception-V3 được thống kê trong Bảng 4.4 sau:

Bảng 4.4. Giá trị trung bình mAP của mô hình Inception-V3

|  |  |  |  |
| --- | --- | --- | --- |
|  | Precision | Recall | F1-score |
| Inception-V3 | 0.855 | 0.812 | 0.83 |

1. **Đánh giá kết quả**

Đối với tập dữ liệu nhận dạng hoa, mô hình Inception-V3 nhận dạng với độ chính xác (precision) và độ bao phủ (recall) trung bình lần lượt là 0.855 và 0.812. Ngoài ra, độ trung bình điều hòa (harmonic mean) F1-score của trung bình precision và recall là 0.83.

Nhìn chung, độ trung bình tổng quan của các giá trị trong mô hình Inception-V3 đạt khá cao

# **CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

1. **Kết quả đạt được**

Nghiên cứu cơ bản đã thực hiện được mục tiêu ban đầu đề ra, đó là tìm hiểu và ứng dụng mô hình học sâu vào bài toán nhận diện hoa, cụ thể là mô hình Inception-V3.

Về tập dữ liệu, thu thập có chọn lọc và tạo tập dữ liệu với đề tài “Nhận dạng hoa sử dụng mô hình học sâu”. Không sử dụng những tập dữ liệu có sẵn, tự thu thập hình ảnh bằng cách tìm kiếm, tự chụp hình ảnh. Tự xử lý, chỉnh sửa, loại bỏ các dữ liệu không đúng định dạng hoặc bị mờ. Dữ liệu sau khi xử lý là 1433 hình ảnh của 25 loại hoa khác nhau.

Về mặt lý thuyết, nghiên cứu đã xây dựng thành công mô hình nhận dạng hoa sử dụng mô hình Inception-V3 và đạt được độ đo F1-score trung bình là 0.83. Độ chính xác trung bình đạt được trên 80% là một kết quả khá tốt so với tập dữ liệu nhỏ.

Xây dựng thành công ứng dụng nhận diện được hình ảnh là của hoa nào và hiển thị được phần trăm độ chính xác tên của hoa đó.

1. **Hướng phát triển**

Thực hiện bổ sung cho tập dữ liệu đa dạng hơn, nhiều lớp hơn và thu tập thêm nhiều hình ảnh hơn cho mỗi lớp dữ liệu.

Xây dựng ứng dụng dễ sử dụng, thân thiện hơn với người dùng, xây dựng thêm các chức năng cho ứng dụng như nhận diện bằng camera bằng cách khi đưa camera vào một bông hoa nào đó thì ứng dụng sẽ quét và hiển thị được tên cũng như thông tin, công dụng của loài hoa đó.

# **TÀI LIỆU THAM KHẢO**

[1] Parvathy, N Vrinda Rao, Shahistha Bai, Naeema Nazer, Prof. Anju. (June 2020). FLOWER RECOGNITION SYSTEM USING CNN. *International Research Journal of Engineering and Technology (IRJET)*, 7(6), 6609-6611.

[2] Ong Zi Yuan, Kah Kien Chye, Huay Wen Kang, Chi Wee Tan. (25- 26 October 2021). A FLOWER RECOGNITION SYSTEM USING DEEP NEURAL NETWORK COUPLED WITH VISUAL GEOMETRY GROUP 19 ARCHITECTURE. *International Conference on Digital Transformation and Applications (ICDXA),* 121-126.

[3] Xiaoling Xia, Cui Xu, Bing Nan. (2 June 2017). Inception-v3 for flower classification. *Computer Science 2017 2nd International Conference on Image, Vision and Computing (ICIVC).*

[4] Mingxing Tan, Quoc V.Le. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Net- works. *ICML.*