Đại học Bách Khoa Hà Nội

Trường Cơ Khí

Khoa Cơ Điện Tử

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A red and white sign with yellow text

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Báo cáo bài tập lớn**

*Môn: Xử lý ảnh*

***Đề tài: Phân loại hoa lan sử dụng mạng neural CNN***

Giảng viên hướng dẫn: Mạc Thị Thoa

Thành viên nhóm:

1. Nguyễn Trung Dũng – 20216065
2. Trần Xuân Thiện – 20216243
3. Thái Khắc Quân – 20216219

Hà Nội, tháng 12/2024

Mục lục

[Phân công nhiệm vụ 4](#_Toc187335052)

[Tóm tắt nội dung bài tập lớn 5](#_Toc187335053)

[Chương I: Đặt vấn đề 7](#_Toc187335054)

[Chương II: Tổng quan về mạng CNN 8](#_Toc187335055)

[2.1 Định nghĩa về mạng neural CNN 8](#_Toc187335056)

[2.2 Kiến trúc của mạng neural CNN 8](#_Toc187335057)

[2.3 Những lớp cơ bản trong mạng neural CNN 10](#_Toc187335058)

[2.4 Cách lựa chọn tham số cho mạng CNN 11](#_Toc187335059)

[Chương III: Xây dựng mạng neural CNN 12](#_Toc187335060)

[Chương IV: Kết quả 14](#_Toc187335061)

[Chương V: Thiết kế giao diện người dùng 15](#_Toc187335062)

[5.1 Phần mềm 15](#_Toc187335063)

[5.2 Giao diện Web 15](#_Toc187335064)

[5.3 Giao diện app android 17](#_Toc187335065)

[Chương VI: Kết luận 21](#_Toc187335066)

[Đánh giá của giảng viên 23](#_Toc187335067)

# Phân công nhiệm vụ

Sau đây là bảng phân công nhiệm vụ của nhóm:

|  |  |  |  |
| --- | --- | --- | --- |
| Người thực hiện | Công việc thực hiện | Mức độ hoàn thành công việc | Thời gian hoàn thiện |
| Nguyễn Trung Dũng | + Tìm kiếm dataset  + Phân loại dataset  + Tối ưu các thông số | Tốt | 3 tuần |
| Trần Xuân Thiện | + Viết code  + Tối ưu các thông số  + Viết web | Tốt | 4 tuần |
| Thái Khắc Quân | + Tìm hiểu sơ bộ về đề tài  + Tìm hiểu về các tools, công cụ được sử dụng trong bài  + Viết báo cáo  + Làm slide | Tốt | 2 tuần |

# Tóm tắt nội dung bài tập lớn

- Đối với bài tập lớn môn Xử lý ảnh, nhóm bọn em tập trung tìm hiểu về các mạng neural nhân tạo, trong đó là trọng tâm là về mạng neural tích chập( Convolution Neural Network)

- Bài tập lớn của nhóm sẽ được chia thành 3 phần chính:

***Phần I : Tìm kiếm và phân loại dataset***

- Đối với hoa lan, nhóm chúng em sẽ chọn đại diện 5 loại hoa lan để làm đầu vào input bao gồm:

* Calopogon tuberosus (phong lan Bắc Mỹ)
* Corallorhiza maculata (lan san hô lốm đốm)
* Cypripedium acaule (lan hài không cuống.)
* Cypripedium parviflorum (lan hài hoa nhỏ)
* Epipactis palustris (lan đầm lầy)

- Các hình ảnh được tìm kiếm trên web: [*https://www.kaggle.com/*](https://www.kaggle.com/)

- Sau khi dữ liệu được thu thập, nó sẽ tiếp tục được chia vào 2 folder là testing ( chiếm 80%) và validation (chiếm 20%) để phục vụ cho việc phân loại và kiểm thử



A screenshot of a computer

Description automatically generated

***Phần II : Viết code để xây dựng mạng neural***

- Các công cụ được sử dụng ở đây bao gồm:

* Framework : Tensoflow/Keras
* Công cụ hỗ trợ lập trình: Jupyter
* Visual Studio Code

- Tóm tắt về cách xử lý dữ liệu:

1. Import các thư viện cần thiết
2. Load ảnh đầu vào mô hình
3. Tăng cường dữ liệu (data augumentation)
4. Xây dựng các layers (Convolution, MaxPooling, Dense)
5. Training trên mô hình vừa xây dựng
6. Đồ thị hoá kết quả thu được
7. Tinh chỉnh lại các thông số để đạt được độ tối ưu hơn cho mô hình
8. Viết web giúp người dùng dễ sử dụng hơn

***Phần III : Viết báo cáo và làm slide thuyết trình***

# Chương I: Đặt vấn đề

- Xử lý ảnh ([tiếng Anh](https://vi.wikipedia.org/wiki/Ti%E1%BA%BFng_Anh): *digital image processing*) hay xử lý ảnh kỹ thuật số là việc sử dụng máy tính số để xử lý các [ảnh kỹ thuật số](https://vi.wikipedia.org/w/index.php?title=%E1%BA%A2nh_k%E1%BB%B9_thu%E1%BA%ADt_s%E1%BB%91&action=edit&redlink=1) thông qua một thuật toán. Là một phân ngành tổng thể của [xử lý tín hiệu số](https://vi.wikipedia.org/wiki/X%E1%BB%AD_l%C3%BD_t%C3%ADn_hi%E1%BB%87u_s%E1%BB%91), xử lý ảnh kỹ thuật số có nhiều lợi thế hơn so với [xử lý ảnh analog](https://vi.wikipedia.org/w/index.php?title=X%E1%BB%AD_l%C3%BD_%E1%BA%A3nh_analog&action=edit&redlink=1). Nó cho phép áp dụng nhiều thuật toán hơn cho dữ liệu đầu vào và có thể tránh được các vấn đề như sự tích tụ [nhiễu](https://vi.wikipedia.org/w/index.php?title=Nhi%E1%BB%85u_t%C3%ADn_hi%E1%BB%87u_s%E1%BB%91&action=edit&redlink=1) và [méo hình](https://vi.wikipedia.org/wiki/%C4%90%E1%BB%99_m%C3%A9o) trong quá trình xử lý. Sự phát triển của xử lý ảnh kỹ thuật số chủ yếu bị ảnh hưởng bởi ba yếu tố:

1. Sự phát triển của máy tính
2. Sự phát triển của toán học (đặc biệt là sáng tạo và cải tiến lý thuyết [toán rời rạc](https://vi.wikipedia.org/wiki/To%C3%A1n_r%E1%BB%9Di_r%E1%BA%A1c));
3. Nhu cầu ứng dụng rộng rãi trong môi trường, nông nghiệp, quân sự, công nghiệp và khoa học y tế ngày càng tăng.

- Ngày nay Xử lý ảnh có một vai trò quan trọng và ngày càng phát triển trong cả công nghiệp và nông nghiệp, giúp nâng cao hiệu quả sản xuất, tối ưu hóa quy trình và giảm chi phí. Dưới đây là các ứng dụng và vai trò cụ thể của xử lý ảnh trong hai lĩnh vực này:

* Kiểm tra chất lượng sản phẩm:
* Nhận diện sản phẩm:
* Theo dõi sự phát triển của cây trồng
* Tưới tiêu thông minh

- Hoa lan (hoặc hoa phong lan) là một trong những loài hoa đẹp và phổ biến nhất trên thế giới, thuộc họ Orchidaceae, một trong những họ thực vật lớn nhất với hơn 25.000 loài và 100.000 giống khác nhau. Hoa lan không chỉ đẹp mắt mà còn mang giá trị kinh tế và văn hóa lớn. Hoa lan được sử dụng phổ biến trong trang trí nhà cửa, đặc biệt là những loài lan có hoa đẹp và dễ chăm sóc. Chúng có thể được trồng trong chậu hoặc dùng để cắm trong các bình hoa. Ngoài ra, hoa lan cũng được chiết xuất để sản xuất các sản phẩm mỹ phẩm, như nước hoa và kem dưỡng da. Một số loài hoa lan, như lan hoàng thảo, được sử dụng trong y học truyền thống của các dân tộc để chữa các bệnh như cảm cúm, đau đầu và các vấn đề về tiêu hóa.

- Để nhận dạng và phân loại hoa lan, có nhiều phương pháp và kỹ thuật khác  
nhau như: logic mờ, giải thuật di truyền, mô hình xác xuất thống kê, mô hình  
mạng nơ ron. Trong đó, mạng nơ-ron tích chập (CNN) đã chứng tỏ được sức mạnh của mình trong việc xử lý hình ảnh và phân loại dữ liệu. Việc áp dụng CNN vào bài toán nhận dạng hoa lan không chỉ mở ra cánh cửa cho việc giải quyết bài toán này một cách hiệu quả mà còn đem lại những tiềm năng lớn trong việc tạo ra các ứng dụng thực tiễn và ứng dụng trong thực tế

- Với sự quan tâm và nỗ lực của nhóm, chúng em đã quyết định tập trung vào đề tài này, hy vọng rằng qua quá trình nghiên cứu và thực hành, chúng em sẽ không chỉ nắm vững kiến thức về mạng CNN mà còn có thể ứng dụng nó vào  
giải quyết bài toán thực tế một cách thành công.

# Chương II: Tổng quan về mạng CNN

## 2.1 Định nghĩa về mạng neural CNN

- CNN là từ viết tắt của cụm Convolutional Neural Network hay là mạng nơ ron tích chập. Đây là mô hình vô cùng tiên tiến được áp dụng nhiều trong lĩnh vực học sâu [Deep learning](https://aptech.fpt.edu.vn/deep-learning.html). Mạng nơ ron này cho phép người dùng xây dựng những hệ thống phân loại và dự đoán với độ chính xác cực cao. Hiện nay, mạng CNN được ứng dụng nhiều hơn trong xử lý ảnh, cụ thể là nhận diện đối tượng trong ảnh.

- Từ “Convolution” trong CNN có nghĩa là một “cửa sổ” sử dụng trượt trên ma trận nhằm lấy được những thông tin chính xác và cần thiết nhất mà không phải chọn đặc trưng (feature). Convolution hay nhân tích chập là cách mà những lớp Convolutional này nhân những phần tử trong ma trận. Sliding Window hay kernel là dạng ma trận có kích thước nhỏ, sử dụng trong nhân tích chập với ma trận hình ảnh.

- Feature trong Convolutional Neural Network: Là những đặc trưng, mạng CNN sẽ so sánh dựa vào từng mảnh và các mảnh như vậy được gọi là feature. Thay vì phải tiến hành khớp các bức ảnh lại với nhau thì mạng nơ ron này sẽ xác định được sự tương đồng thông qua tìm kiếm thô những đặc trưng khớp với nhau bằng hai hình ảnh tốt hơn. Một feature là mộ hình ảnh dạng mini (những mảng 2 chiều nhỏ). Những feature này đều tương ứng với một khía cạnh nào đó của hình ảnh và chúng có thể khớp lại được với nhau.

## 2.2 Kiến trúc của mạng neural CNN

- Mạng CNN là tập hợp những Convolutional layer xếp chồng lên nhau, đồng thời mạng sử dụng những hàm như ReLU và Tanh để kích hoạt các trọng số trong các node. Các lớp này sau khi qua các hàm activation sẽ có trọng số trong những node và có thể tạo ra những thông tin trừu tượng hơn đến với các lớp kế tiếp trong mạng.

- Mạng này có tính kết hợp cả tính bất biến. Tức là, nếu cùng một đối tượng mà sử dụng chiếu theo các góc độ khác nhau thì sẽ có ảnh hưởng đến độ chính xác. Với dịch chuyển, co giãn hay quay ma trận ảnh thì lớp Pooling sẽ được dùng để hỗ trợ làm bất biến các tính chất này. Chính vì vậy mà mạng nơ ron này sẽ đưa ra những kết quả có độ chính xác tương ứng với từng mô hình.

- Trong đó, lớp Pooling sẽ có khả năng tạo tính bất biến với phép dịch chuyển, co giãn và quay. Còn tính kết hợp cục bộ sẽ cho thấy những cấp độ biểu diễn, dữ liệu từ thấp đến cao với mức trừu tượng thông qua Convolution từ filter. Mạng CNN có những lớp liên kết nhau dựa vào cơ chế Convolution.

- Các lớp tiếp theo sẽ là kết quả từ những lớp trước đó, vì vậy mà bạn sẽ có những liên kết cục bộ phù hợp nhất. Trong quá trình huấn luyện mạng, mạng nơ ron này sẽ tự học hỏi những giá trị thông qua filter layer dựa theo cách thức mà bạn thực hiện.

- Cấu trúc cơ bản của một mô hình mạng CNN thường bao gồm 3 phần chính bao gồm:

* **Trường cục bộ/ Local receptive field:** Lớp này sử dụng để tách lọc dữ liệu, thông tin hình ảnh để từ đó có thể lựa chọn các vùng có giá trị sử dụng hiệu quả cao nhất.
* **Trọng số chia sẻ/ Shared weights and bias:** Lớp này hỗ trợ làm giảm các tham số đến mức tối thiểu trong mạng CNN. Trong từng lớp convolution sẽ chứa các feature map riêng và từng feature thì sẽ có khả năng phát hiện một vài feature trong hình ảnh.
* **Lớp tổng hợp/ Pooling layer:** Đây là lớp cuối cùng và sử dụng để làm đơn giản các thông tin output. Tức là, sau khi tính toán xong và quét qua các layer trong mạng thì pooling layer sẽ được dùng để lược bỏ các thông tin không hữu ích. Từ đó cho ra kết quả theo kỳ vọng người dùng.

A diagram of a diagram of a pooling and pooling

Description automatically generated with medium confidence

## 2.3 Những lớp cơ bản trong mạng neural CNN

**Convolutional layer**

- Lớp này là phần quan trọng nhất của toàn mạng CNN, nó có nhiệm vụ thực thi các tính toán. Các yếu tố quan trọng trong lớp Convolutional là: padding, stride, feature map và filter map.

* Sử dụng filter để áp dụng vào các vùng của ma trận hình ảnh. Các filter map là các ma trận 3 chiều, bên trong đó là những tham số và chúng được gọi là parameters.
* Stride tức là bạn dịch chuyển filter map theo từng pixel dựa vào các giá trị từ trái qua phải.
* Padding: Thường, giá trị viền xung quanh của ma trận hình ảnh sẽ được gán các giá trị 0 để có thể tiến hành nhân tích chập mà không làm giảm kích thước ma trận ảnh ban đầu.
* Feature map: Biểu diễn kết quả sau mỗi lần feature map quét qua ma trận ảnh đầu vào. Sau mỗi lần quét thì lớp Convolutional  sẽ tiến hành tính toán.

A diagram of a mathematical equation

Description automatically generated with medium confidence

**Relu Layer**

- Lớp ReLU này là hàm kích hoạt trong mạng CNN, được gọi là activation function. Nó có tác dụng mô phỏng những nơ ron có tỷ lệ truyền xung qua axon. Các hàm activation khác như Leaky, Sigmoid, Leaky, Maxout,.. tuy nhiên hiện nay, hàm ReLU được sử dụng phổ biến và thông dụng nhất.

- Hàm này được sử dụng cho những yêu cầu huấn luyện mạng nơ ron với những ưu điểm nổi bật điển hình là hỗ trợ tính toán nhanh hơn. Trong quá trình dùng hàm ReLU, bạn cần chú ý đến việc tùy chỉnh những learning rate và dead unit. Những lớp ReLU được dùng sau khi filter map được tính và áp dụng ReLU lên các giá trị của filter map.

**Pooling layer**

- Khi ma trận ảnh đầu vào có kích thước quá lớn, các lớp Pooling layer sẽ được đặt vào giữa những lớp Convolutional để làm giảm những parameters. Hiện, hai loại lớp Pooling được sử dụng phổ biến là Max pooling và Average.

A diagram of a pool

Description automatically generated

**Fully connected layer**

- Đây là lớp có nhiệm vụ đưa ra kết quả sau khi hai lớp Convolutional và Pooling đả nhận được ảnh truyền. Khi này, ta sẽ thu được một model đọc được thông tin của ảnh. Để có thể liên kế chúng cũng như cho nhiều đầu ra hơn ta sẽ sử dụng Fully connected layer.

- Ngoài ra, nếu lớp này có dữ liệu hình ảnh thì lớp sẽ chuyển chúng thành các much chưa được phân chia chất lượng để tìm ra ảnh có chất lượng cao nhất.

## 2.4 Cách lựa chọn tham số cho mạng CNN

- Để chọn tham số phù hợp nhất cho mạng Convolutional Neural Network thì bạn cần chú ý đến những yếu tố như: filter size, số convolution, pooling size và việc train – test.

* **Lớp Convolution:** Số lượng lớp này càng nhiều thì sẽ giúp cải thiện được hoạt động của chương trình. Sử dụng những lớp với số lượng lớn thì khả năng hạn chế các tác động các tốt. Thông thường, chỉ sau khoảng 3 đến 4 lớp bạn sẽ đạt được kết quả như kỳ vọng
* **Filter size:** Kích thước thường chọn là ma trận 3×3 hoặc ma trận 5×5.
* **Pooling size:** Với những hình ảnh thông thường, bạn nên chọn ma trận pooling kích thước 2×2. Với những ảnh kích thước lớn thì nên chọn ma trận kích thước 3×3.
* **Train test**: Thực hiện nhiều lần để cho ra kết quả tốt nhất

# Chương III: Xây dựng mạng neural CNN

- Để xây dựng 1 mạng neural CNN cơ bản thì sẽ cần trải qua các bước sau đây:

1. Import các thư viện cần dùng

A close-up of a computer screen

Description automatically generated

1. Load ảnh đầu vào

A screenshot of a computer

Description automatically generated

1. Data Augumentation

A white background with a black border

Description automatically generated with medium confidence

A collage of a flower

Description automatically generated

1. Xây dựng các layers

A white background with a black and white flag

Description automatically generated with medium confidence

1. Training cho mạng neural và kiểm tra hiệu quả



A screenshot of a computer

Description automatically generated

# Chương IV: Kết quả

- Kết quả nhận được :

A graph showing the number of epops

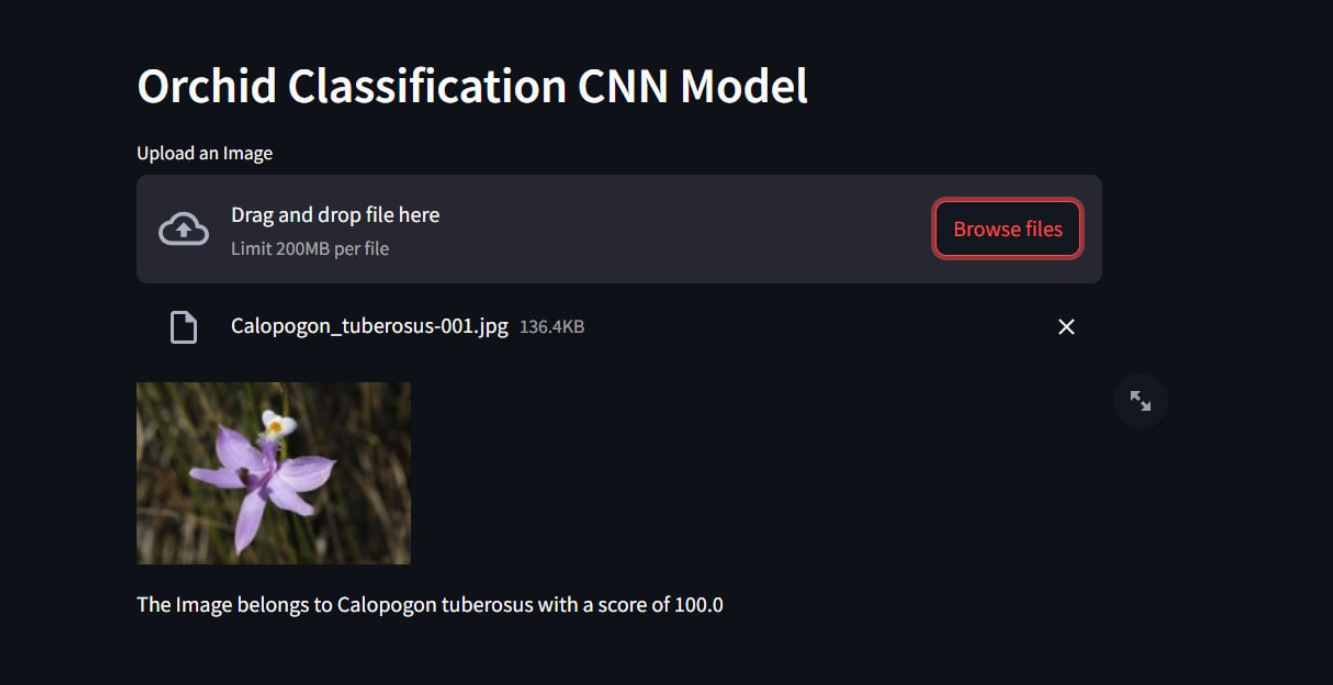
Description automatically generated

80 epochs

A graph showing the number of epoxy

Description automatically generated

150 epochs



Giao diện web

# Chương V: Thiết kế giao diện người dùng

## 5.1 Phần mềm

Khi đã training xong mô hình, cần đưa nó vào phần mềm giao diện để người dùng có thể dễ dàng thao tác. Sử dụng Visual studio code và Android studio để thiết kế giao diện web và app android.

## 5.2 Giao diện Web

Lưu lại model dưới file .h5:

A white rectangular object with green text

Description automatically generated

Sử dụng thư viện streamlit để hiển thị web ở localhost, import các thư viện cần thiết khác:

A screen shot of a computer

Description automatically generated

Load model và tạo một mảng chứa tên các loại hoa:

A screenshot of a computer program

Description automatically generated

Tạo phần load ảnh vào model và xử lý kết quả đầu ra:

A screen shot of a computer program

Description automatically generated

Cuối cùng là giao diện web hoàn chỉnh:

A screenshot of a computer

Description automatically generated

## 5.3 Giao diện app android

Đầu tiên, thiết kế giao diện khởi tạo, bao gồm các thông tin sinh viên, môn học, giảng viên, giao diện này sẽ là giao diện hiện ra khi khởi động ứng dụng:

A screenshot of a computer

Description automatically generated

Tiếp theo, thiết kế giao diện chính, chức năng gồm các nút load ảnh từ thư viện hoặc nút chụp ảnh, ảnh sẽ được hiện vào ImageView, khi nhận diện xong sẽ hiển thị vào TextView, nút trở về để về màn hình ban đầu:

A screenshot of a computer

Description automatically generated

Tiến hành code backEnd của ứng dụng:

* Thiết lập nút bấm A screenshot of a computer code

  Description automatically generated

A screenshot of a computer program

Description automatically generated

* Hàm xử lý lấy ảnh từ thư viện, camera:

A screen shot of a computer code

Description automatically generated

* Hàm xử lý khi ảnh nạp vào model

A screenshot of a computer program

Description automatically generated

Sau khi build file apk, cài app, ta hoàn được giao diện như sau

A screen shot of a phone

Description automatically generatedA screenshot of a video game

Description automatically generated

# Chương VI: Kết luận

- Trong bài tập lớn này, nhóm chúng em đã nghiên cứu và xây dựng một mạng neural tích chập (CNN) để phân loại các loại hoa dựa trên hình ảnh đầu vào. Qua quá trình thực hiện, nhóm đã đạt được những kết quả và bài học như sau:

* **Hiệu quả của mạng CNN:**

Mạng CNN được thiết kế và huấn luyện đã chứng minh khả năng phân loại các loại hoa với độ chính xác tương đối cao (khoảng 90%). Điều này khẳng định sức mạnh của mạng CNN trong xử lý và phân tích dữ liệu hình ảnh.

* **Thử nghiệm và tối ưu:**

Chúng em đã thực hiện nhiều thử nghiệm với các kiến trúc khác nhau của CNN, thay đổi các siêu tham số như kích thước lớp chập, số lượng tầng, và hàm kích hoạt. Kết quả cho thấy rằng việc tinh chỉnh siêu tham số và tiền xử lý dữ liệu là yếu tố quan trọng ảnh hưởng đến hiệu suất của mô hình.

* **Khó khăn gặp phải:**

Một số dữ liệu hình ảnh trong tập huấn luyện và kiểm tra có chất lượng thấp (mờ, nhiễu), dẫn đến giảm độ chính xác trong một số trường hợp.

Tập dữ liệu bị mất cân bằng giữa các lớp hoa cũng gây ra hiện tượng thiên vị trong quá trình huấn luyện.

Gặp tình trạng overfitting

* **Định hướng phát triển:**

Mở rộng tập dữ liệu với số lượng hình ảnh lớn hơn và đa dạng hơn để tăng độ chính xác.

Sử dụng các phương pháp tăng cường dữ liệu (data augmentation) để cải thiện tính tổng quát của mô hình.

Thử nghiệm với các mô hình CNN tiên tiến hơn như ResNet, VGGNet, YOLO hoặc ứng dụng Transfer Learning.

* **Tổng kết:**

Qua bài tập lớn này, nhóm không chỉ hiểu sâu hơn về nguyên lý hoạt động của mạng CNN mà còn nắm bắt được quy trình xây dựng, huấn luyện và tối ưu hóa mô hình trong thực tế. Đây là nền tảng quan trọng cho những nghiên cứu và ứng dụng tiếp theo trong lĩnh vực thị giác máy tính.

# Đánh giá của giảng viên

*Nhận xét của giảng viên hướng dẫn :*

Hà Nội, ngày tháng năm 2025

Giảng viên hướng dẫn

(Ký và ghi rõ họ tên)