**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA ĐA PHƯƠNG TIỆN**



**COMPUTER VISION**

How to prepare data for CV problem

| **Giảng viên:** | Phạm Văn Sự |
| --- | --- |
| **Nhóm:** | 12 |
| **Thành viên:** | Nguyễn Quốc Đạt - B20DCPT053  Nguyễn Văn Tuấn - B20DCPT182  Nguyễn Hồng Nhung - B20DCPT150 |

HÀ NỘI, THÁNG 11/2023

[**MỤC LỤC**](#_2eayp5e05243) **2**

[**I. Khảo sát về các bộ dữ liệu:**](#_vru64hexqznk) **3**

[MNIST](#_zf4cwyuat2zk) 3

[(Modified National Institute of Standards and Technology)](#_8ep4mmaon69d) 3

[Coco (Common Objects in Context)](#_njaijzakvu3z) 3

[EMNIST (Extended MNIST)](#_xeutnmz8pfed) 3

[Caltech 101](#_dhhbeivji02h) 3

[Caltech 256](#_c6oew2dgatbt) 4

[CIFAR-10 (Canadian Institute for Advanced Research - 10)](#_qul741927z0t) 4

[CIFAR-100 (Canadian Institute for Advanced Research - 100)](#_ueusj0vh70hh) 4

[Image Net](#_l4pvdiqu2d8w) 5

[Fashion-MNIST](#_c5v6ihydqul1) 5

[Oxford 102 Flowers Dataset](#_f14ha3l6xdp3) 5

[PASCAL VOC (Visual Object Classes)](#_cdej2yviu2u0) 5

[Stanford Dogs Dataset](#_6bokyos7v1j0) 6

[Stanford Cars Dataset](#_w6d5869j3kra) 6

[**II. Công cụ gán nhãn LabelImg**](#_8ll5vo1vkm08) **6**

[2. Cách cài đặt:](#_tx857s6zxc67) 6

[3. Thực hành:](#_wourw8s6vx34) 7

[**III. Chuẩn bị dữ liệu**](#_jhjua3n3zjln) **7**

[1. Download dữ liệu:](#_todv7nrw8o3n) 7

[- Dữ liệu được down bằng tay từ các nguồn ảnh trên mạng với định dạng là file .jpeg hoặc .jpg.](#_pppp6yofvdpo) 7

[2. Chuẩn bị dữ liệu:](#_juw1t16x4l8k) 7

[3. Train:](#_1obga3z3y70h) 7

# **Khảo sát về các bộ dữ liệu:**

| **STT** | **Tên bộ dữ liệu** | **Tác giả** | **Các đặc điểm chung** | **Download** |
| --- | --- | --- | --- | --- |
| 1 | MNIST(Modified National Institute of Standards and Technology) | Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner | * Về các chữ số viết tay. * Bao gồm 70000 ảnh. * 10 lớp (từ 0 đến 9). * Đặc điểm: ảnh xám, kích thước 28x28 pixels. * Năm công bố 1998. | * Url: <https://www.kaggle.com/datasets/hojjatk/mnist-dataset/data> * Được tích hợp trong thư viện scikit-learn |
| 2 | Coco (Common Objects in Context) | Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Larry Zitnick , và Piotr Dollar | * Về các đối tượng, thực thể trong tình huống thực tế. * Hơn 330k ảnh (>200k ảnh được gán nhãn). * Năm công bố: 2014 | * Url: <https://cocodataset.org/#download> |
| 3 | EMNIST (Extended MNIST) | Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre van Schaik | * Về các chữ số và chữ cái viết tay, mở rộng về nhiều biến thể. * Được chia làm 6 loại. * Năm ra mắt: 2017 | * Url: [ICNS | EMNIST (western sydney.edu.au)](https://www.westernsydney.edu.au/icns/resources/reproducible_research3/publication_support_materials2/emnist) |
| 4 | Caltech 101 | [Li Fei-Fei](https://ieeexplore.ieee.org/author/37087189080), [R. Fergus](https://ieeexplore.ieee.org/author/37419045500),  [P. Perona](https://ieeexplore.ieee.org/author/37282902900) | * Về các đối tượng khác nhau như người, xe, động vật …. * Gồm 9000 hình ảnh khác nhau được phân loại làm 101 lớp, mỗi lớp từ 40-800 bức ảnh. * Đặc điểm: Kích thước mỗi ảnh 300x200 pixels * Ra mắt năm 2003 | * Url: <https://data.caltech.edu/records/mzrjq-6wc02> * Url: [caltech101 | TensorFlow Datasets](https://www.tensorflow.org/datasets/catalog/caltech101) |
| 5 | Caltech 256 | Li Fei-Fei, Marco Andreetto, and Marc 'Aurelio Ranzato | * Về các đối tượng, thực thể. * Bao gồm 30607 ảnh. * 257 lớp đối tượng. * Đặc điểm: ảnh màu, kích thước không cố định. * Năm công bố: không xác định. | * Url:<https://www.kaggle.com/datasets/jessicali9530/caltech256> |
| 6 | CIFAR-10 (Canadian Institute for Advanced Research - 10) | Alex Krizhevsky, Vinod Nair, Geoffrey Hinton | * Về các đối tượng thường thấy như ô tô, máy bay, thú cưng, … * Bao gồm 60000 ảnh. * 10 lớp đối tượng. * Đặc điểm: Ảnh màu, kích thước 32x32 pixels. * Năm công bố: 2009. | * Url:<https://www.kaggle.com/competitions/cifar-10/data> |
| 7 | CIFAR-100 (Canadian Institute for Advanced Research - 100) | Alex Krizhevsky, Vinod Nair, Geoffrey Hinton | * Về các đối tượng thường thấy như ô tô, máy bay, thú cưng, … * Bao gồm 60000 ảnh, 600 ảnh mỗi lớp. * 100 lớp đối tượng chia thành 5 lớp siêu cấp. * Đặc điểm: Ảnh màu, kích thước 32x32 pixels. * Năm công bố: 2009. | * Url: <https://www.kaggle.com/datasets/fedesoriano/cifar100> |
| 8 | Image Net | [Jia Deng](https://ieeexplore.ieee.org/author/37085544091), [Wei Dong](https://ieeexplore.ieee.org/author/38277671200), [Richard Socher](https://ieeexplore.ieee.org/author/37547163200), [Li-Jia Li](https://ieeexplore.ieee.org/author/38273260200), [Kai Li](https://ieeexplore.ieee.org/author/37087178742), [Li Fei-Fei](https://ieeexplore.ieee.org/author/37087189080) | * Về các đối tượng, vật thể, … * Bao gồm hơn 14 triệu ảnh. * Hơn 20000 lớp. * Đặc điểm: Ảnh màu, kích thước không cố định. * Năm công bố 2009. | * Url: <https://www.image-net.org/challenges/LSVRC/index.php> * Url: <https://www.kaggle.com/c/imagenet-object-localization-challenge/overview/description> |
| 9 | Fashion-MNIST | Không rõ tác giả | * Về các loại quần áo, giày dép thời trang khác nhau. * Gồm 70000 ảnh với 60000 ảnh huấn luyện và 10000 thử nghiệm. * Đặc điểm: Kích thước 28x28 pixel. * Năm công bố: 2017 | * Url: <https://www.kaggle.com/datasets/zalando-research/fashionmnist> |
| 10 | Oxford 102 Flowers Dataset | [Maria-Elena Nilsback](http://www.robots.ox.ac.uk/~men/) and [Andrew Zisserman](http://www.robots.ox.ac.uk/~az/) | * Về 102 loài hoa phổ biến ở nước Anh. Mỗi lớp từ 40-258 ảnh. * Gồm 16185 bức ảnh. * Ra mắt năm 2008. | * Url: [Visual Geometry Group - University of Oxford](https://www.robots.ox.ac.uk/~vgg/data/flowers/102/) |
| 11 | PASCAL VOC (Visual Object Classes) | Mark Everingham, Luc van Gool, Chris Williams, John Winn, Andrew Zisserman, Yusuf Aytar, Ali Eslami, Alexander Sorokin | * Về các đối tượng người, xe hơi, động vật… * Chia thành ba tập hợp con: 1.464 hình ảnh để huấn luyện, 1.449 hình ảnh để xác thực và một bộ thử nghiệm riêng. * Năm công bố: 2012 | * Url: <https://www.kaggle.com/datasets/gopalbhattrai/pascal-voc-2012-dataset> |
| 12 | Stanford Dogs Dataset | [Aditya Khosla](http://people.csail.mit.edu/khosla/), [Nityananda Jayadevaprakash](http://vision.stanford.edu/aditya86/ImageNetDogs/main.html#), [Bangpeng Yao](http://ai.stanford.edu/~bangpeng/), [Li Fei-Fei](http://vision.stanford.edu/) | * Về hơn 120 loài chó trên khắp thế giới. * Gồm 20580 ảnh * Ra mắt năm 2011. | * Url: <http://vision.stanford.edu/aditya86/ImageNetDogs/main.html> |
| 13 | Stanford Cars Dataset | Jonathan Krause | * Về hơn 196 lớp xe trên thế giới * Hơn 16185 bức ảnh chia thành 2 phần bằng nhau với 8144 ảnh huấn luyện và 8041 cho phần kiểm tra. * Ra mắt năm 2013 | * Url: <https://www.kaggle.com/datasets/jessicali9530/stanford-cars-dataset> |

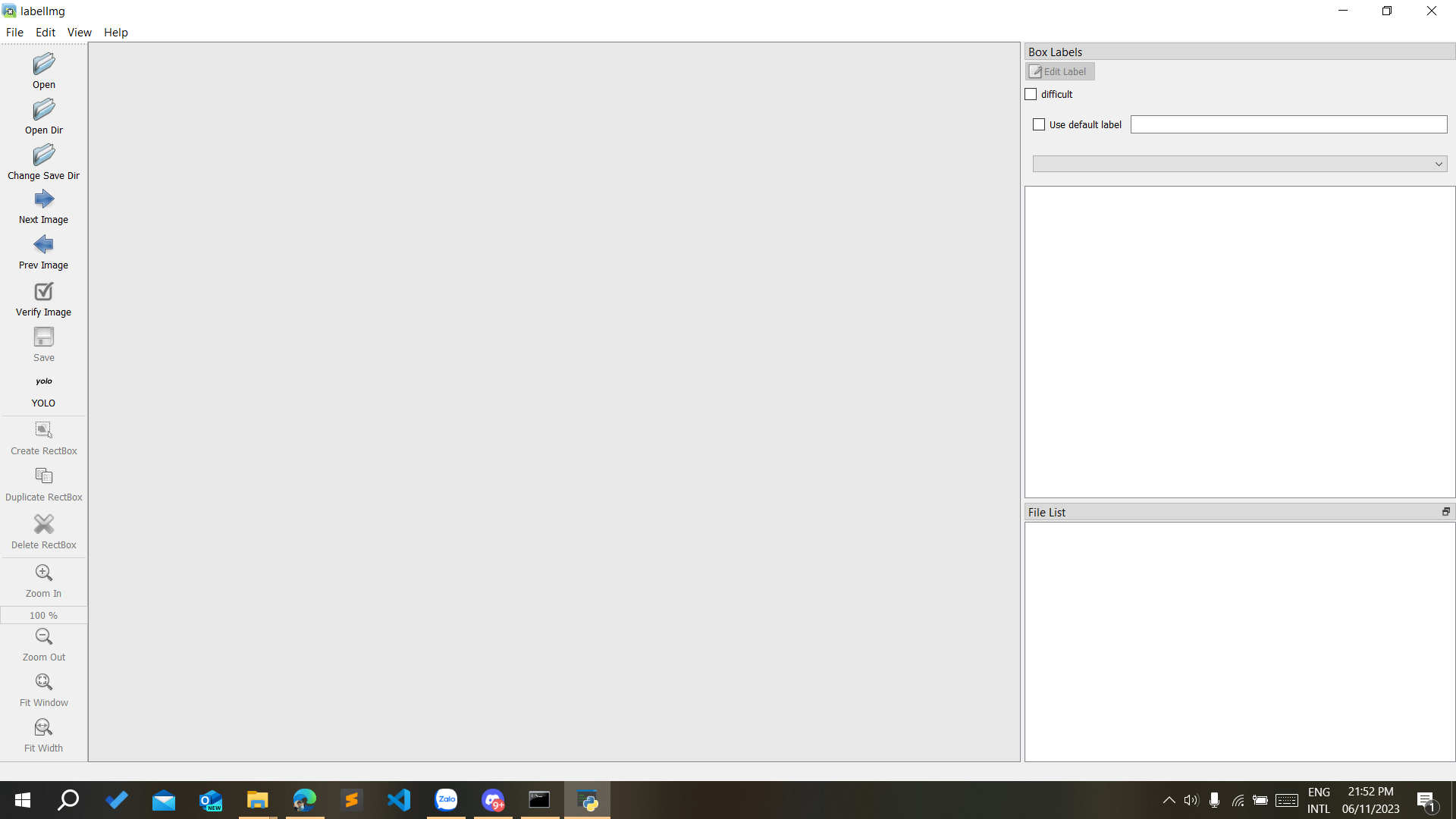
# **Công cụ gán nhãn LabelImg**

1. Giới thiệu về LabelImg:

* Gán nhãn hình ảnh là bước đầu và là bước quan trọng để phát hiện đối tượng trong ảnh. Quá trình gán nhãn mất rất nhiều thời gian, nhưng nếu bạn càng gán được nhiều nhãn thì mô hình huấn luyện của bạn lại càng chính xác.
* LabelImg là một công cụ gán nhãn đồ họa. Nó được viết bằng Python và sử dụng thư viện Qt cho giao diện.
* Nhãn được lưu dưới dạng tệp xml đối với Pascal VOC - định dạng được sử dụng bởi ImageNet. Bên cạnh đó nó cũng được hỗ trợ bởi YOLO.

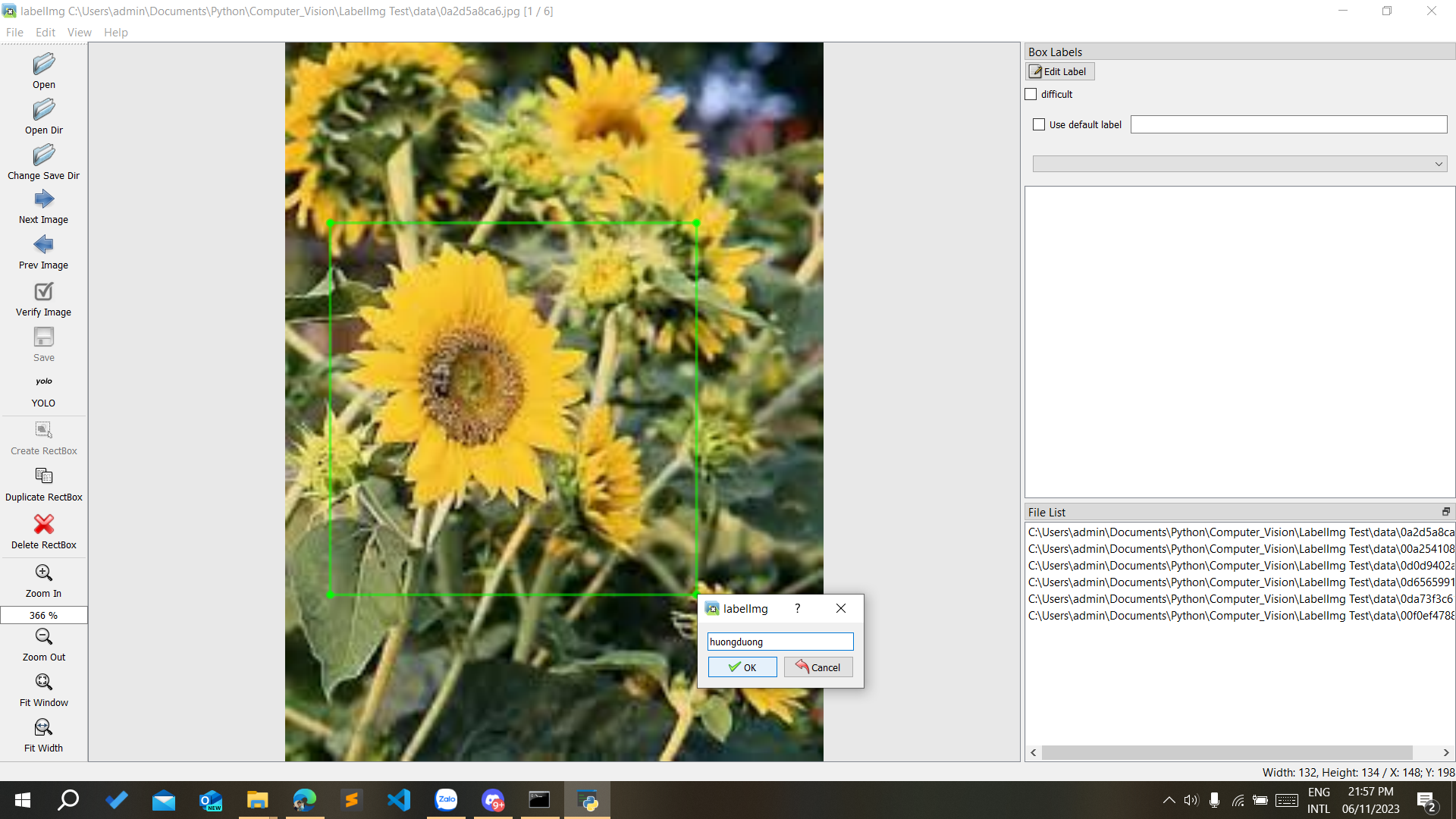
## Cách cài đặt:

* Link hướng dẫn: <https://github.com/HumanSignal/labelImg/tree/master#steps-pascalvoc>
* Trước tiên cài đặt thư viện lxml, pyqt5, labelImg, Sử dụng lệnh pip install lxml pyqt5 labelImg.
* Vào thư mục lưu data, gõ cmd. Sau đó giao diện ứng dụng hiện lên.



* Open là để mở 1 file ảnh.
* Open Dir để mở thư mục ảnh.
* Change Save Dir để thay đổi thư mục lưu định dạng các file log ra.
* Next/Previous để chuyển ảnh tiếp theo hoặc mở ảnh trước đó.
* Khi gán nhãn cho ảnh cần phải chọn chế độ gán nhãn. Ở đây có 3 chế độ, bao gồm PascalVoc, Yolo, CreateXML. Cụ thể 3 chế độ này như thế nào thì trên links gắn phía trên đã nói rõ.
* Khi lưu với dạng Yolo thì file class.txt sẽ lưu tên các nhãn lớp.

## Thực hành:



# **Chuẩn bị dữ liệu**

Tổng quan về bộ dữ liệu: là tập hợp khoảng hơn 700 ảnh về 5 loài hoa ở Việt Nam.

## Download dữ liệu

## Dữ liệu được down bằng tay từ các nguồn ảnh trên mạng với định dạng là file .jpeg hoặc .jpg.

* Dữ liệu cũng được chia thành 5 tập hợp với 5 loại: hoa dã quỳ, hoa đào, hoa hồng, hoa hướng dương, hoa ly.
* Bộ dữ liệu đa dạng về kích thước (trên 125 ảnh với mỗi lớp).

## Chuẩn bị dữ liệu

* Loại bỏ một số file không hợp lệ hoặc không đúng định dạng.

Script minh họa:

# read data

X, y = [], []

for category in categories:

folderpath = path + '/' + category

for img in os.listdir(folderpath):

img\_path = folderpath + '/' + img

if img\_path.endswith('.jpg') or img\_path.endswith('.jpeg'):

X.append(img\_path)

y.append(category)

* Chia tập dữ liệu thành 3 tập bao gồm train, valid, test.

Script minh họa:

train\_df, dummy\_df = train\_test\_split(df, train\_size= 0.8, shuffle= *True*, random\_state= 123)

# valid and test dataframe

valid\_df, test\_df = train\_test\_split(dummy\_df, train\_size= 0.6, shuffle= *True*, random\_state= 123)

* Scale lại ảnh cho đồng bộ về kích thước

Script minh họa:

batch\_size = 16

img\_size = (224, 224)

channels = 3

img\_shape = (img\_size[0], img\_size[1], channels)

tr\_gen = ImageDataGenerator()

ts\_gen = ImageDataGenerator()

train\_gen = tr\_gen.flow\_from\_dataframe( train\_df, x\_col= 'filepaths', y\_col= 'labels', target\_size= img\_size, class\_mode= 'categorical',

color\_mode= 'rgb', shuffle= *True*, batch\_size= batch\_size)

valid\_gen = ts\_gen.flow\_from\_dataframe( valid\_df, x\_col= 'filepaths', y\_col= 'labels', target\_size= img\_size, class\_mode= 'categorical',

color\_mode= 'rgb', shuffle= *True*, batch\_size= batch\_size)

test\_gen = ts\_gen.flow\_from\_dataframe( test\_df, x\_col= 'filepaths', y\_col= 'labels', target\_size= img\_size, class\_mode= 'categorical',

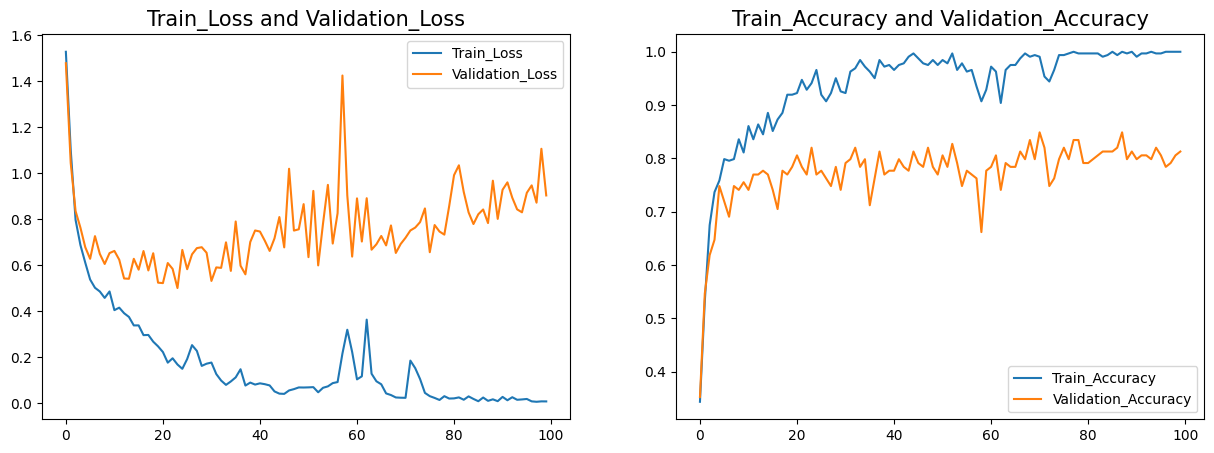
color\_mode= 'rgb', shuffle= *False*, batch\_size= batch\_size)

## Train:

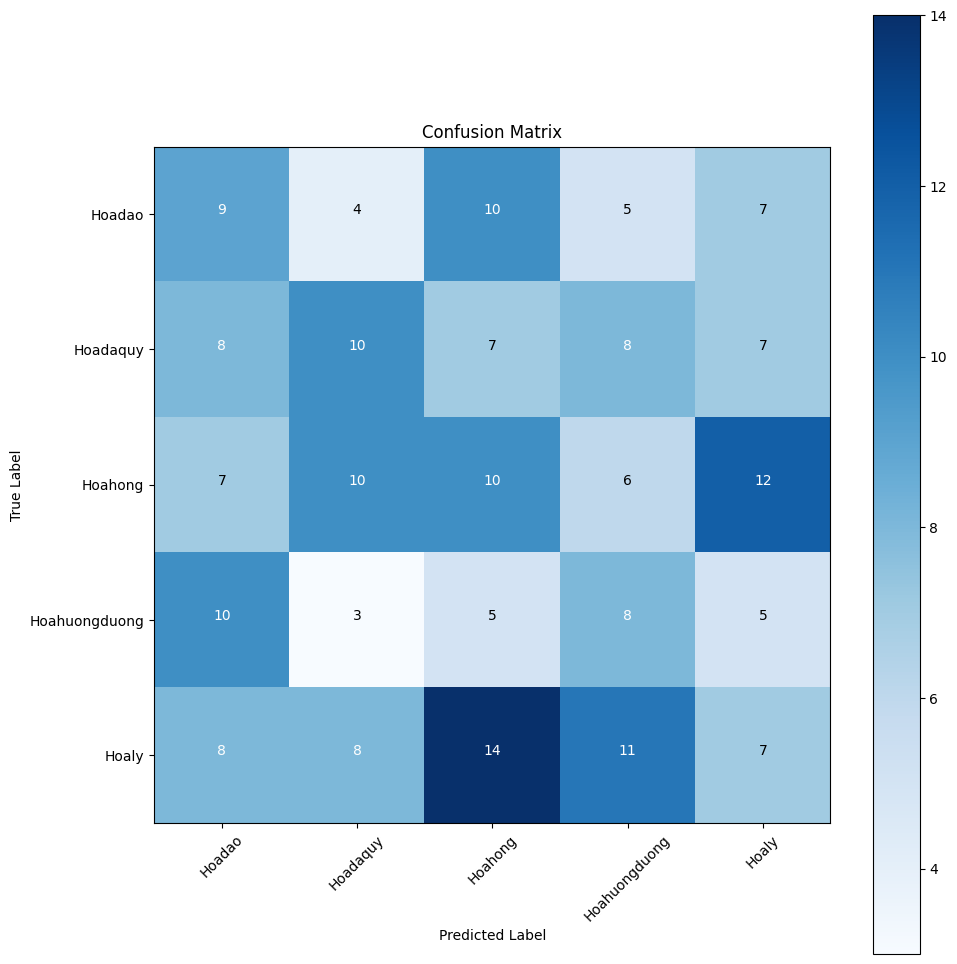
* Model sử dụng train dựa trên basic CNN:

src train: <https://colab.research.google.com/drive/1tDFqyuHJx9yBz5uKfdjsrHZeKlOGH-fG#scrollTo=yrOKyzpzUZFs>

Đánh giá, kiểm thử dựa trên: train\_loss, val\_loss, train\_accuracy và val\_accuracy.



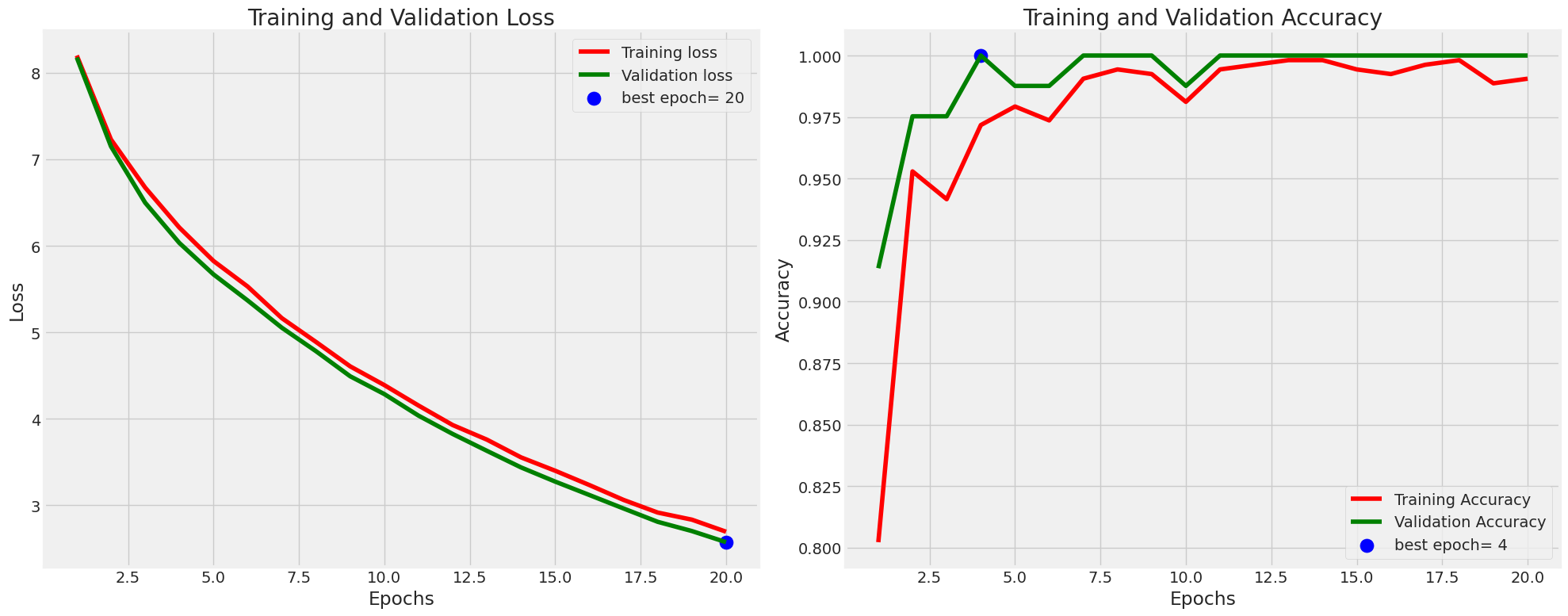
Confusion matrix:



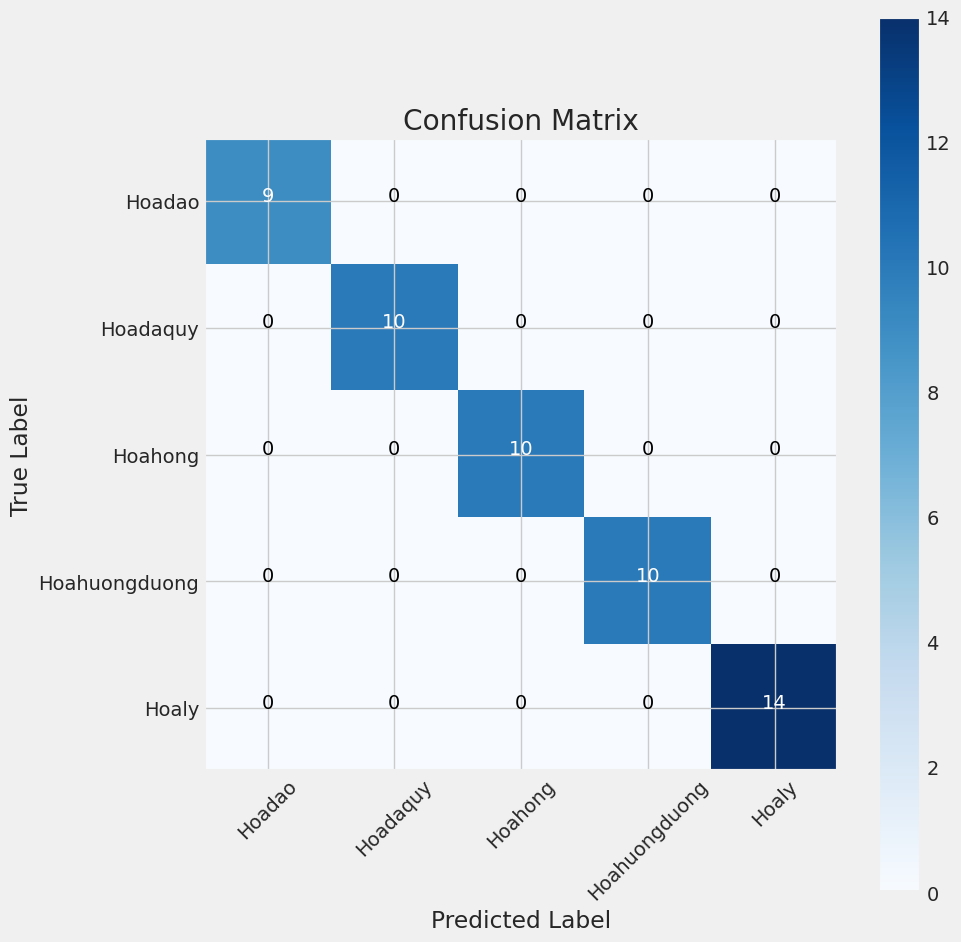
* Model sử dụng Efficient NetB3:

src train: <https://colab.research.google.com/drive/1tDFqyuHJx9yBz5uKfdjsrHZeKlOGH-fG#scrollTo=yrOKyzpzUZFs>

Đánh giá, kiểm thử dựa trên: train\_loss, val\_loss, train\_accuracy và val\_accuracy.



Confusion Matrix



Classes Report: 