Fruit Quality Classification

[Group 2 - B2 DS]

Hoang Kim Huong - 22BA13158
Nguyen Thi Ngoc Lan - 22BA13186
Doan Ngoc Linh - 22BA13190
Le Sy Han - 23BI14150
Le Hoang Dat - 23BI14087

Table of Contents

01. Introduction

1. Background & Content
2. Project Objectives
3. Scope

Background & Content

Why Fruit Freshness Classification Matters

- Ensures consumer health
- Reduces post-harvest food waste
- Improves supply chain efficiency

Problem

- Current manual inspection methods: slow and error-prone

=> Automated, accurated, scalabled classification system

Project Objectives

- Build an automated system using image analysis
- Improve quality control processes
- Reduce post-harvest losses
- Support food safety efforts
- Contribute to smart agriculture and modern supply chain management

Scope

- Uses Convolutional Neural Network (CNN)
- Focus on 3 fruits: apples, bananas, oranges
- Simple classification: fresh or rotten
- Narrow scope => high accuracy
- Foundation for future expansion (more fruits, more detailed freshness levels)

02. Dataset

Data Description

Data Preprocessing

Image Collection

6 folders with 300 images per folder, ensuring a balanced representation across all classes

Image Loading

Using OpenCV (cv2.imread)
Failed to load => Skip and show a warning message

Data Preprocessing

Resizing

A uniform dimension 256 x 256 pixels to standardize input size, to ensure compatibility with CNN

Normalization

Scale to the [0.0, 1.0] range

Data Preprocessing

Label Extraction

Based on the folder names, each label combine fruit type and freshness condition

Data Storage

Storing in a NumPy array with the shape (1800, 256, 256, 3)

Encode categorical into numeric form

Label encoding

Encoding ‘freshapples’, ‘freshbanana’, ‘freshoranges’, ‘rottenapples’, ‘rottenbanana’, ‘rottenoranges’ to 0 to 5

One Hot Encoding

‘freshapples’ to [1;0;0;0;0]
‘freshbanana’ to [0;1;0;0;0]
...

Shuffle & Split the dataset

- Shuffle the ordered dataset into the unordered dataset
- Split the dataset into 3 subsets
  + Initial Split: A training + validation set 90% and a testing set 10%
  + Secondary Split: A training set 80% and a validation set 20%

03.
Related Work

K-Means Clustering Algorithm

K-Means Clustering Algorithm

04. Algorithm & Model

Convolutional Neural Network

(CNN)

4.1. Algorithm of CNN

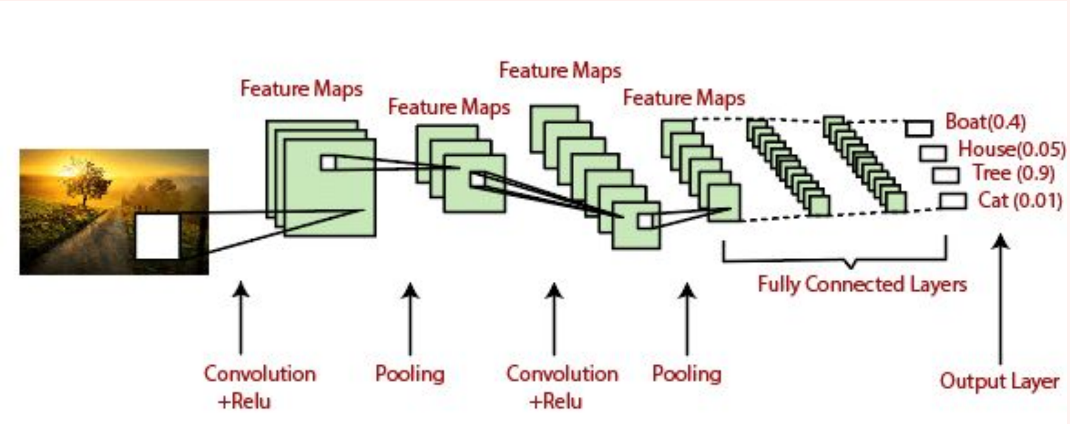
Convolutional Neural Network (CNN): specialized deep learning model

- Autonomous feature extraction
- Translation-invariant characteristics
- Identify patterns and features despite variations in position, orientation, scale, or translation

4.2. Components of CNN

4 main components:

1. Convolutional layers
2. Activation function
3. Pooling layers
4. Fully-connected layers



4.3. Model's Architecture

Contains: 3 convolutional blocks & the fully connected layers

Layer Type	Hyper-parameter
Conv2D	32 filters of size 3x3, ReLU
MaxPooling2D	3x3 pool, stride (2,2)
Dropout	Rate = 0.3

[1st Convolution Block]

Layer Type	Hyper-parameter
Conv2D	64 filters of size 3x3, ReLU
MaxPooling2D	3x3 pool, stride (2,2)
Dropout	Rate = 0.3

[2nd Convolutional Block]

Layer Type	Hyper-parameter
Conv2D	128 filters of size 2x2, ReLU
MaxPooling2D	2x2 pool, stride (2,2)
Dropout	Rate = 0.3

[3rd Convolutional Block]

4.3. Model's Architecture

Contains: 3 convolutional blocks & the fully connected layers

[Fully connected layers]

Layer Type	Hyper-parameter
Flatten	
Dense	256 units, ReLU
Dropout	Rate = 0.3
Output Dense	6 units, softmax activation

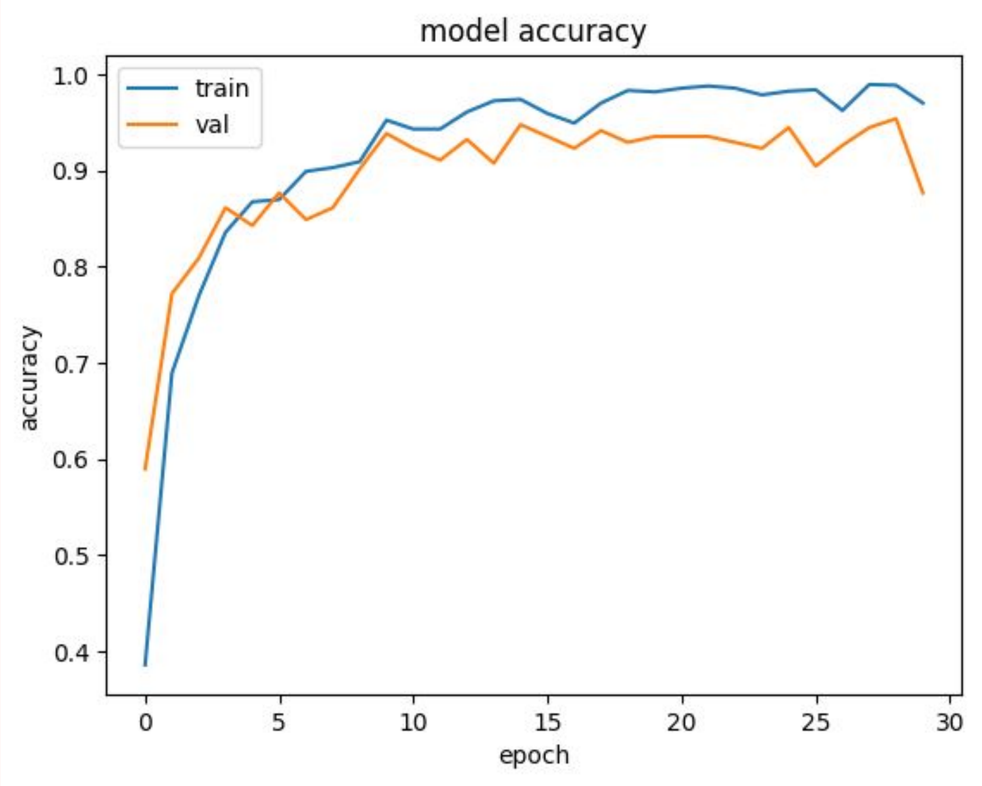
05. Result & Evaluation

Training vs Validation Performance

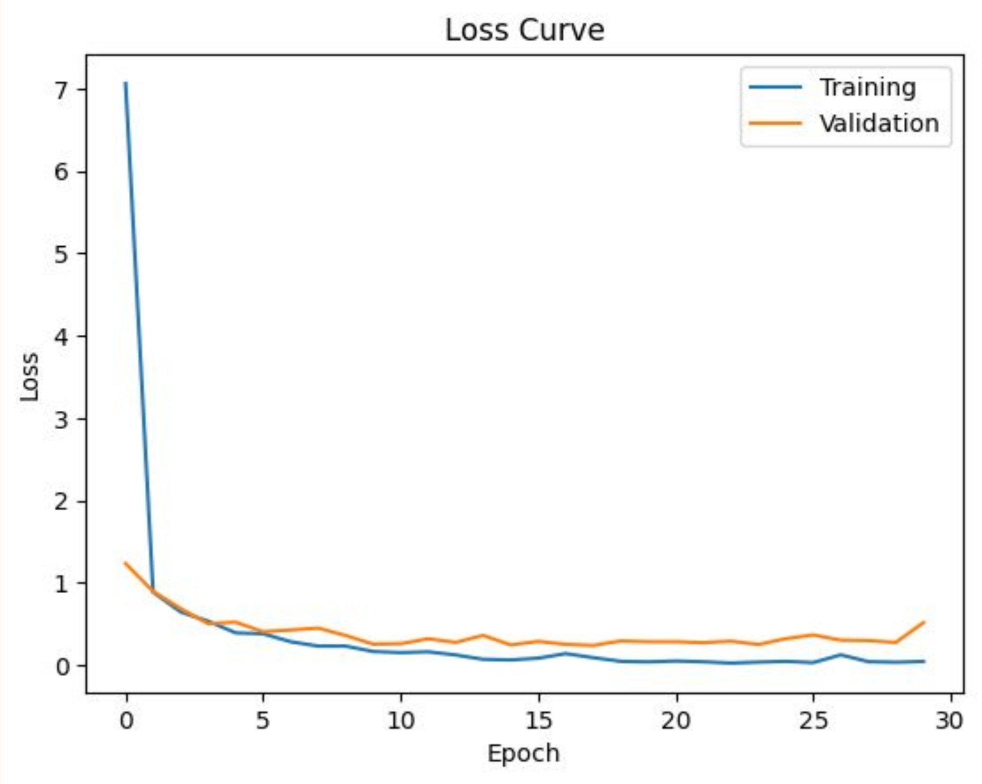
Confusion matrix

Classification metrics

5.1.Training vs Validation Performance



5.1.Training vs Validation Performance



5.2.Confusion Matrix

		Actual					
		Class 0	Class 1	Class 2	Class 3	Class 4	Class 5
Predicted	Class 0	30	0	0	2	0	0
	Class 1	0	32	0	0	1	0
	Class 2	0	1	36	0	0	2
	Class 3	3	0	0	22	1	0
	Class 4	0	1	0	0	18	0
	Class 5	1	1	3	1	0	25

Table 1: Confusion Matrix of the Fruit Quality Classification Model

5.3.Classification Metrics

Class	Precision	Recall	F1-score	Support
0	0.88	0.94	0.91	32
1	0.91	0.97	0.94	33
2	0.92	0.92	0.92	39
3	0.88	0.85	0.86	26
4	0.90	0.95	0.92	19
5	0.93	0.81	0.86	31
Accuracy			0.91	180
Macro average	0.90	0.91	0.90	180
Weighted average	0.91	0.91	0.90	180

Table 2: Classification Report of the Fruit Quality Classification model

06.

Discussion & Conclusion

Discussion & Conclusion

CNNs are potential in automating fruit quality classification tasks.

Misclassifications between fresh and rotten samples of the same fruit may due to subtle visual differences or inconsistent lighting conditions in the dataset.

The relatively small dataset size may restrict the model’s robustness across broader conditions.

Discussion & Conclusion

Future work:

- Increasing dataset diversity.
- Enhancing the model’s ability to evaluate and classify.
- Optimizing for speed and performance.
- Deploying & testing the model in real-world environments.

Thank you!