Machine Learning - ha4

Weili Diao Student ID: 21127071

Program: Big Data Techonology Email: wdiaoaa@connect.ust.hk

1 Introduction

1.1 Project Objective

The task of this assignment is to classify the given image using six models, including AlexNet, VGG16, ResNet50, InceptionV3, DenseNet121 and MobileNetV2, then summarize and analyze the results, discuss the similarities and differences of the results, and analyze the reasons.

This report first introduces the preparation of the experiment (image preprocessing and environment configuration), then displays the experimental results, and finally discusses the experimental results, especially for some specific ones.

1.2 Preparation

1.2.1 Images

Before conducting experiments, the read images need to be preprocessed. The steps are as follows:

1. Convert images to RGB format.
2. Resize and crop images to fit model input needs.
3. Convert the images to tensor and convert its pixel values into a range of [0, 1].
4. Normalize the images using the following transform:

```
normalize = transforms.Normalize(mean=[0.485, 0.456, 0.406],
                                 std=[0.229, 0.224, 0.225])
```

1.2.2 Environment

Python version: 3.10.12
Operating System: Linux 6.1.85+
PyTorch version: 2.5.0+cu121

2 Experiment

2.1 Results

These tables present the outputs of models:

Image	AlexNet	VGG16	ResNet50
1	sunscreen(0.2120)	golfcart(0.7300)	barrow(0.3523)
2	pop_bottle(0.1321)	vacuum(0.2545)	water_bottle(0.3702)
3	library(0.7063)	library(0.9486)	library(0.7547)
4	car_mirror(0.9935)	car_mirror(0.9903)	car_mirror(0.9859)
5	oscilloscope(0.3488)	desktop_computer(0.3030)	desk(0.2679)
6	zebra(0.9999)	zebra(1.0000)	zebra(1.0000)
7	school_bus(1.0000)	school_bus(1.0000)	school_bus(1.0000)
8	pillow(0.9878)	pillow(1.0000)	pillow(1.0000)
9	fireboat(1.0000)	fireboat(1.0000)	fireboat(1.0000)
10	carousel(1.0000)	carousel(1.0000)	carousel(1.0000)

Image	InceptionV3	DenseNet121	MobileNetV2
1	golfcart(0.9997)	golfcart(0.2394)	golfcart(0.6980)
2	water_bottle(0.9219)	water_bottle(0.8169)	tennis_ball(0.1077)
3	library(0.8807)	library(0.9107)	library(0.8926)
4	car_mirror(0.9984)	car_mirror(0.9805)	car_mirror(0.9837)
5	laptop(0.6180)	notebook(0.2593)	desktop_computer(0.2458)
6	zebra(1.0000)	zebra(1.0000)	zebra(1.0000)
7	school_bus(1.0000)	school_bus(1.0000)	school_bus(1.0000)
8	pillow(0.9999)	pillow(1.0000)	pillow(1.0000)
9	fireboat(1.0000)	fireboat(1.0000)	fireboat(1.0000)
10	carousel(0.9962)	carousel(1.0000)	carousel(1.0000)

Note: The numbers in parentheses represent the probabilities of classification.

2.2 Discussion

2.2.1 Analysis of Results

In this experiment, the accuracy of each model is shown in the table below:

Model	Accuracy
AlexNet	0.7
VGG16	0.9
ResNet50	0.8
InceptionV3	0.9
DenseNet121	0.9
MobileNetV2	0.9

The classification accuracy of each image is as follows:

Image	1	2	3	4	5	6	7	8	9	10
Classification accuracy	0.67	0.5	1	1	0.33	1	1	1	1	1

Overall, the model performs the worst on the fifth image, only with an accuracy of 0.33. Besides, we can notice that although the model MobileNetV2 predicts that the probability of

the fifth image being a desktop computer is only 0.2458, the classification result is correct. On the contrary, the model InceptionV3 predicted that the probability of the fifth image being a laptop is up to 0.6180, but the result is wrong. In this regard, we additionally output the complete classification results of the fifth image by these two models.

Classification(top 6) of the fifth image by InceptionV3:

Classification	Probability
laptop	0.6180
desktop_computer	0.1487
screen	0.0795
mouse	0.0520
desk	0.0434
oscilloscope	0.0143

Classification(top 6) of the fifth image by MobileNetV2:

Classification	Probability
desktop_computer	0.2458
desk	0.1617
mouse	0.1526
screen	0.1203
notebook	0.0995
monitor	0.0899

2.2.2 Analysis of Reasons

Comparing the classification accuracy of the six models, we can see that the accuracy of the AlexNet model is relatively low. But for parameters, this model has the second largest number of parameters among the six models. But in consideration of structure, this model is shallower, has relatively fewer layers, and has larger convolution kernels, which affects its accuracy in this task.

For the first, second and fifth images, the models' classification resulta are not ideal, but they can accurately classify other seven images. We can notice that these images have one thing in common: they have more noise.

Take the fifth image as an example. There is a notebook in the image, but the DenseNet121 model was interfered by the notebook and did not get the expected results. Similarly, as shown in the Classification(top 6) of the fifth image by MobileNetV2 table above, all six possible outcomes appear in the image, making it impossible for the model to distinguish which one is the focus of the image. When there are multiple objects in the image, the model will not be able to accurately obtain features or obtain confusing features.

In addition, some pictures are classified into similar objects. As shown in the Classification(top 6) of the fifth image by InceptionV3 table above, desktop computer is in its classification results, but the probability is smaller than laptop. Because the two are similar, but for this model, there may be a lack of enough training samples to extract features that can distinguish the two.