**Vision AI intern assignment**

**Exercise 1 – Image classification**

1. **Mục tiêu bài toán.**

Xây dựng mô hình có thể phân loại hình ảnh chó và mèo từ hình ảnh đầu vào.

Input: ảnh

Output: Chó or mèo

1. **Dữ liệu huấn luyện.**

Dữ liệu huấn luyện được thu thập trên Kaggle: [**Cats and Dogs Classification Dataset**](https://www.kaggle.com/datasets/bhavikjikadara/dog-and-cat-classification-dataset)

Dữ liệu bao gồm 12500 hình ảnh chó và 12500 hình ảnh mèo được chia vào 2 thư mục Dog và Cat.

Sau khi thu thập dữ liệu tiến hành phân chia theo 3 tập train, val, test theo thỉ lệ: 7:2:1.

Cấu trúc dataset như bên dưới.

Ảnh có chứa văn bản, biểu đồ, ảnh chụp màn hình, Phông chữ

Nội dung do AI tạo ra có thể không chính xác.

Xử lý hình ảnh trước khi train:

* Resize kích thước ảnh về cùng 1 kích cở 224x224
* Lật ngẫu nhiêu ảnh theo chiều ngang với xác suất 0.5 giúp tăng sự đa dạng của data.
* Xoay ảnh ngẫu nhiêu khoảng 10 độ. -> tăng sự đa dạng của data.
* Chuyển image sang Tensor để đưa vào model của Pytorch.

1. **Kiến trúc mô hình.**

Trong bài toán phân loại ảnh chó và mèo, tôi thử nghiệm nhiều mô hình khác nhau để so sánh hiệu năng và khả năng tổng quát hóa. Các mô hình được triển khai bằng PyTorch.

Dưới đây là các model mà tôi dùng:

* CNN

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm

Nội dung do AI tạo ra có thể không chính xác.

* ResNet50:

Sử dụng ResNet pretrained trên ImageNet, thay đổi fully connected layer cuối.

* MobileNetV2
* EfficientNetB0

1. **Pineline .**

Ảnh có chứa văn bản, Phông chữ, ảnh chụp màn hình, hàng

Nội dung do AI tạo ra có thể không chính xác.

1. **Kết quả.**

|  |  |
| --- | --- |
| Model | Accuracy |
| CNN |  |
| ResNet50 |  |
| MobileNetV2 |  |
| EfficientNetB0 |  |
|  |  |

1. **Vấn đề tồn tại và hướng giải quyết.**

Lúc đầu train bằng CNN bị overfitting nhẹ. Lý do là do lúc đó chỉ tiền xử lý bằng resize hình ảnh. Sau đó, đã tiến hành tăng tính đa dạng hình ảnh bằng cách thêm xử lý xoay và lật ảnh. Kết quả là đã giúp model ổn định hơn trước đó. Kết quả đánh giá cũng tốt hơn.

**Exercise 2 – Text to speech**