TRƯỜNG ĐẠI HỌC HỌC VĂN LANG

**KHOA CÔNG NGHỆ THÔNG TIN**

Ảnh có chứa biểu tượng, Phông chữ, Đồ họa, văn bản

Mô tả được tạo tự động

**BÁO CÁO ĐỒ ÁN MÔN HỌC HK241**

**HỌC MÁY VÀ ỨNG DỤNG**

**ĐỀ TÀI : NHẬN DẠNG HÌNH ẢNH BẰNG CÁC**

**MÔ HÌNH HUẤN LUYỆNVỚI DỮ LIỆU CIFAR-10**

**SVTH : TẠ GIA BẢO**

**MSSV : 2274802010050**

**GVHD : HUỲNH THÁI HỌC**

TP. Hồ Chí Minh – năm 2024

**MỤC LỤC**

[1. MÔ TẢ DỮ LIỆU 2](#_Toc182605766)

[2. MỤC ĐÍCH 3](#_Toc182605767)

[3. CÁC MÔ HÌNH HUẤN LUYỆN 4](#_Toc182605768)

[3.1 Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) 4](#_Toc182605769)

[3.2 ResNet (Residual Network) 5](#_Toc182605770)

[3.3 VGGNet 5](#_Toc182605771)

[4. KẾT QUẢ 5](#_Toc182605772)

[5. ĐÁNH GIÁ, NHẬN XÉT: 9](#_Toc182605773)

[5.1 Đánh giá 9](#_Toc182605774)

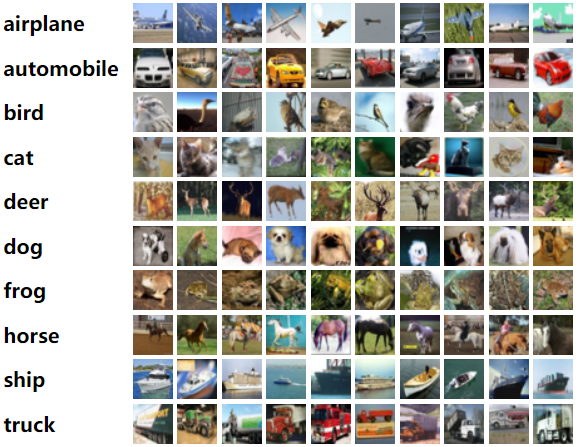
[5.2 Nhận xét 10](#_Toc182605775)

# MÔ TẢ DỮ LIỆU

Tập dữ liệu CIFAR-10 được phát triển bởi nhóm nghiên cứu tại **Canadian Institute for Advanced Research (CIFAR)**. Tập dữ liệu này được công bố lần đầu tiên vào năm 2009 bởi **Alex Krizhevsky**, Geoffrey Hinton và các cộng sự.

[CIFAR-10](http://www.cs.toronto.edu/~kriz/cifar.html)  là một tập dữ liệu thị giác máy tính đã được thiết lập được sử dụng để nhận dạng đối tượng. Đây là một tập hợp con của [tập dữ liệu 80 triệu hình ảnh nhỏ](http://groups.csail.mit.edu/vision/TinyImages/) và bao gồm 60.000 hình ảnh màu 32x32 chứa một trong 10 lớp đối tượng, với 6000 hình ảnh cho mỗi lớp.

Sau đây là các lớp trong bộ dữ liệu, cũng như 10 ảnh ngẫu nhiên từ mỗi lớp:



*Hình 1: Các lớp trong dữ liệu CIFAR-10*

Mỗi lớp có 6,000 hình ảnh, trong đó:

* Tập huấn luyện: 5,000 hình ảnh cho mỗi lớp (tổng cộng 50,000 hình ảnh).
* Tập kiểm tra: 1,000 hình ảnh cho mỗi lớp (tổng cộng 10,000 hình ảnh).

Tỷ lệ giữa tập huấn luyện và tập kiểm tra:

* Tập huấn luyện: 50,000 / 60,000 = 83.33%
* Tập kiểm tra: 10,000 / 60,000 = 16.67%

# MỤC ĐÍCH

Mục tiêu của tập dữ liệu CIFAR-10 là huấn luyện các mô hình học máy, đặc biệt là các mạng nơ-ron tích chập (CNN), để phân loại hình ảnh thành 10 lớp khác nhau, ví dụ như dự đoán các đối tượng như máy bay, xe hơi, chim, mèo, v.v. Mỗi hình ảnh trong CIFAR-10 có thể được gán một nhãn từ 1 trong 10 lớp này.

Input đầu vào: Các hình ảnh màu 32x32 pixels, là đầu vào cho mô hình.

Nhãn (Output): Nhãn phân loại của mỗi hình ảnh, xác định đối tượng mà hình ảnh đại diện (ví dụ, "dog", "airplane", "cat", v.v.).

Ý nghĩa: CIFAR-10 giúp nghiên cứu và phát triển các mô hình phân loại hình ảnh tự động. Nó cung cấp một nền tảng cho việc thử nghiệm và đánh giá các kỹ thuật học sâu, đặc biệt là các mạng nơ-ron tích chập (CNNs), trong việc xử lý và phân loại các hình ảnh nhỏ với độ phân giải thấp. Tập dữ liệu này giúp hiểu và cải thiện khả năng nhận diện hình ảnh trong các ứng dụng như nhận diện đối tượng, xe tự lái, và các hệ thống trí tuệ nhân tạo khác.

# CÁC MÔ HÌNH HUẤN LUYỆN

## Mạng nơ-ron tích chập (Convolutional Neural Network - CNN)

CNN có khả năng nhận dạng và học các đặc trưng không gian trong hình ảnh, chẳng hạn như các cạnh, kết cấu và hình dạng. Điều này giúp CNN rất mạnh mẽ trong việc phân loại hình ảnh.

Lợi ích của CNN:

* Convolutional layers giúp phát hiện các đặc trưng (features) như cạnh, góc, và kết cấu mà các mô hình khác không thể nhận dạng hiệu quả.
* Pooling layers giúp giảm độ phức tạp và làm mờ chi tiết không quan trọng, giúp mô hình trở nên ổn định hơn với các biến đổi nhỏ trong ảnh (như thay đổi kích thước hoặc vị trí).
* Rất phù hợp với dữ liệu hình ảnh: CNN có thể học được các đặc trưng hình ảnh ở nhiều mức độ khác nhau và có khả năng tổng quát tốt.
* CNN có thể đạt được độ chính xác cao và được sử dụng rộng rãi trong các bài toán như phân loại hình ảnh, nhận diện đối tượng, và nhận diện văn bản.

## ResNet (Residual Network)

ResNet sử dụng một kiến trúc đặc biệt gọi là **residual block**. Thay vì học trực tiếp một phép biến đổi đầu ra, mỗi block học sự khác biệt (residual) giữa đầu vào và đầu ra. Điều này giúp giảm thiểu vấn đề gradient biến mất (vanishing gradient) khi đào tạo các mạng sâu.

Đặc điểm của ResNet là các skip connections, nơi thông tin từ các tầng trước được truyền trực tiếp tới các tầng sau. Điều này giúp tránh việc mất mát thông tin và cải thiện khả năng huấn luyện mạng nơ-ron sâu.

ResNet có thể được xây dựng với nhiều tầng khác nhau, ví dụ như ResNet-34, ResNet-50, ResNet-101, ResNet-110, với số lượng tầng ngày càng tăng. Các mô hình này đều sử dụng residual blocks để cải thiện hiệu quả huấn luyện.

## VGGNet

VGGNet là một kiến trúc mạng nơ-ron tích chập sâu (CNN) nổi tiếng, được phát triển bởi nhóm nghiên cứu tại Đại học Oxford. Được giới thiệu lần đầu tiên trong cuộc thi ImageNet Large Scale Visual Recognition Challenge (ILSVRC) năm 2014, VGGNet nổi bật với cấu trúc đơn giản nhưng hiệu quả, và là một trong những mô hình đầu tiên đạt được hiệu suất cao bằng cách tăng độ sâu của mạng.

VGGNet là mô hình CNN sâu với kiến trúc đơn giản và hiệu quả, gồm các lớp tích chập 3x3 xếp chồng để đạt độ sâu lớn, với hai phiên bản phổ biến là VGG-16 và VGG-19. Dù yêu cầu cao về bộ nhớ và tính toán, VGGNet mang lại hiệu suất xuất sắc trong các bài toán thị giác máy tính, từ phân loại đến nhận diện ảnh.

# KẾT QUẢ

Với mô hình CNN thì sau khi huấn luyện 30 vòng có kết quả như sau

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

Mô tả được tạo tự động

Ảnh có chứa phương tiện, Phương tiện đường bộ, bánh xe, ngoài trời

Mô tả được tạo tự động Nhận dạng hình ảnh qua 2 tấm ảnh:

Ảnh có chứa con mèo, động vật có vú, mèo nhà, Mèo cỡ nhỏ tới cỡ trung

Mô tả được tạo tự động

Độ tin cậy 95% Độ tin cậy 37%

Với mô hình ResNet sau khi huấn luyện 30 vòng có kết quả như sau:

Ảnh có chứa văn bản, ảnh chụp màn hình

Mô tả được tạo tự động

Ảnh có chứa phương tiện, Phương tiện đường bộ, bánh xe, ngoài trời

Mô tả được tạo tự độngNhận dạng hình ảnh qua 2 tấm ảnh:

Ảnh có chứa con mèo, động vật có vú, mèo nhà, Mèo cỡ nhỏ tới cỡ trung

Mô tả được tạo tự động

Độ tin cậy 93% Độ tin cậy 76%

Với mô hình VGG sau khi huấn luyện 30 vòng có kết quả như sau:

Ảnh có chứa văn bản, ảnh chụp màn hình, thiết kế, tác phẩm nghệ thuật

Mô tả được tạo tự động

Nhận dạng hình ảnh qua 2 tấm ảnh:

Ảnh có chứa con mèo, động vật có vú, Mèo cỡ nhỏ tới cỡ trung, Họ Mèo

Mô tả được tạo tự độngẢnh có chứa phương tiện, Phương tiện đường bộ, bánh xe, xe thể thao

Mô tả được tạo tự động

Độ tin cậy 10% và sai nhận dạng Độ tin cậy 10% và sai nhận dạng

# ĐÁNH GIÁ, NHẬN XÉT:

## 5.1 Đánh giá

Mô hình CNN:

* Mô hình CNN có một mức độ mất mát (loss) khá cao so với các mô hình khác, nhưng vẫn duy trì một độ chính xác hợp lý (59%). Điều này có thể chỉ ra rằng mô hình đang học được một số đặc trưng nhưng vẫn chưa hoàn toàn tối ưu.
* Có thể cần tinh chỉnh thêm các tham số (learning rate, batch size) hoặc cải thiện dữ liệu đầu vào để cải thiện độ chính xác.

Mô hình ResNet:

* ResNet đạt được kết quả tốt nhất trong ba mô hình, với loss thấp hơn (0.83) và độ chính xác khá cao (73%).
* ResNet có khả năng học các đặc trưng phức tạp nhờ vào các lớp residual (các lớp kết nối tắt), giúp giảm hiện tượng vanishing gradient và cải thiện hiệu suất học.
* Đây là mô hình đáng xem xét để triển khai thực tế, vì nó cung cấp một sự cân bằng tốt giữa độ chính xác và mất mát.

Mô hình VGG:

* Kết quả của VGG là rất thấp, với độ mất mát (loss) rất cao và độ chính xác chỉ có 10%. Điều này cho thấy mô hình không học được nhiều từ dữ liệu huấn luyện và đang gặp vấn đề nghiêm trọng.
* Nguyên nhân có thể là do kiến trúc VGG có quá nhiều lớp và chưa được tinh chỉnh đúng mức cho bài toán cụ thể, hoặc có thể là do vấn đề về dữ liệu (ví dụ, dữ liệu không đủ hoặc bị thiếu sót, hoặc có sự mất cân đối giữa các lớp).
* Nếu dữ liệu không được tiền xử lý tốt hoặc không đủ để huấn luyện một mô hình phức tạp như VGG, kết quả sẽ rất tệ.

## 5.2 Nhận xét

ResNet là mô hình có kết quả tốt nhất, với độ chính xác cao và mất mát thấp. Đây là lựa chọn hợp lý nếu muốn có một mô hình ổn định và hiệu quả.

CNN cho thấy hiệu suất ổn định nhưng cần cải thiện để tối ưu hơn nữa, có thể thử tinh chỉnh thêm các siêu tham số hoặc áp dụng các kỹ thuật cải thiện độ chính xác như data augmentation hoặc dropout.

VGG có kết quả tồi tệ nhất và có thể cần cải thiện trong việc tiền xử lý dữ liệu hoặc điều chỉnh lại kiến trúc mô hình. Ta có thể thử giảm số lượng lớp hoặc thay đổi các tham số học.

**Link github code : https://github.com/tajs1510/Machine-Learning/tree/main/Project**