TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN



**ĐỒ ÁN HỌC SÂU**

**PHẦN MỀM NHẬN DẠNG 5 ĐỐI TƯỢNG**

GIẢNG VIÊN HƯỚNG DẪN: ThS. Tôn Quang Toại

SINH VIÊN THỰC HIỆN:

Bùi Tuấn Đạt – 21DH113218

Cao Thảo Vân – 21DH113129

Ngô Anh Tuấn – 21DH113091

Phạm Gia Huy – 21DH113678

**TP. HỒ CHÍ MINH – 04/2024**

TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN



**ĐỒ ÁN HỌC SÂU**

**PHẦN MỀM NHẬN DẠNG 5 ĐỐI TƯỢNG**

GIẢNG VIÊN HƯỚNG DẪN: ThS. Tôn Quang Toại

SINH VIÊN THỰC HIỆN:

Bùi Tuấn Đạt – 21DH113218

Cao Thảo Vân – 21DH113129

Ngô Anh Tuấn – 21DH113091

Phạm Gia Huy – 21DH113678

**TP. HỒ CHÍ MINH – 04/2024**

PHIẾU CHẤM ĐIỂM

Họ tên sinh viên 1 (SV1): Mã SV:

Họ tên sinh viên 2 (SV2): Mã SV:

Họ tên sinh viên 3 (SV3): Mã SV:

Họ tên sinh viên 4 (SV4): Mã SV:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CLO** | **Nội dung/Chuẩn đầu ra** | **ĐIỂM CỦA SV 1** | **ĐIỂM CỦA SV 2** | **ĐIỂM CỦA SV 3** | **ĐIỂM CỦA SV 4** |
| 1 | Xây dựng và huấn luyện mạng neuron  (Bài toán phân lớp ảnh, huấn luyện mô hình phân lớp…) |  |  |  |  |
| 2 | Tinh chỉnh mô hình và thuật toán huấn luyện  (Bài toán phân đoạn, tìm gốc quay, phát hiện vị trí, …) |  |  |  |  |
| 3 | Vận dụng mạng neuron tích chập và mạng hồi quy  (Triển khai mô hình trên web, desktop, mobile, …) |  |  |  |  |
| 4 | Có khả năng giải quyết một số vấn đề thực tế  (Thu thập dữ liệu, xử lý dữ liệu, …) |  |  |  |  |
| 5 | Có năng lực trình bày giải pháp kỹ thuật  (Thuyết trình, trình bày báo cáo, …) |  |  |  |  |
| Tổng | |  |  |  |  |

Họ tên GV 1: Ký tên:

Họ tên GV 2: Ký tên:

TÓM TẮT ĐỒ ÁN

Đề tài này nhằm phát triển phần mềm AI có khả năng tự động nhận dạng 5 đối tượng cụ thể trong ảnh và xác định chính xác vị trí của chúng.

Input: Ảnh có thể chứa 0 đến nhiều đối tượng trong số 5 đối tượng đã chọn.

Output: Ảnh kèm theo thông tin về vị trí (dạng bounding boxes) và nhãn của các đối tượng được nhận dạng.

Mục tiêu là tạo ra một công cụ hữu ích cho việc phân loại và phát hiện đối tượng tự động trong các lĩnh vực như an ninh, quảng cáo, hoặc nghiên cứu khoa học.

MỤC LỤC

[TÓM TẮT ĐỒ ÁN I](#_Toc163018484)

[MỤC LỤC II](#_Toc163018485)

[Chương 1. Giới thiệu bài toán 1](#_Toc163018486)

[1.1 Câu hỏi nghiên cứu 1](#_Toc163018487)

[1.2 Giới hạn nghiên cứu 2](#_Toc163018488)

[1.3 Bố cục đồ án 2](#_Toc163018489)

[Chương 2. Giải pháp đề xuất 4](#_Toc163018490)

[2.1 Phân tích dữ liệu 4](#_Toc163018491)

[2.2 Mô hình 5](#_Toc163018492)

[2.2.1 Mô hình MLP 5](#_Toc163018493)

[2.2.2 Mô hình CNN 7](#_Toc163018494)

[2.2.3 Mô hình ResNet 50 8](#_Toc163018495)

[2.2.4 Mô hình DenseNet 9](#_Toc163018496)

[Chương 3. Thực nghiệm 11](#_Toc163018497)

[1. Mô hình MLP: 11](#_Toc163018498)

[2. Mô hình CNN: 11](#_Toc163018499)

[3. Mô hình Resnet50: 12](#_Toc163018500)

[4. Mô hình DenseNet: 13](#_Toc163018501)

[KẾT LUẬN 15](#_Toc163018502)

[TÀI LIỆU THAM KHẢO 16](#_Toc163018503)

1. Giới thiệu bài toán
   1. Câu hỏi nghiên cứu

- **Phát biểu nội dung bài toán:** Bài toán đặt ra là phát triển một mô hình máy học sâu có khả năng nhận diện 5 đối tượng cụ thể trong các bức ảnh. Bên cạnh việc nhận diện, mô hình cũng cần xác định chính xác vị trí của từng đối tượng trong ảnh thông qua việc vẽ bounding box xung quanh đối tượng.

- **Mô tả Input của bài toán:** Input của bài toán là các bức ảnh màu hoặc ảnh xám, có thể có định dạng khác nhau như JPG, PNG. Ảnh có thể chứa một hoặc nhiều trong số 5 đối tượng cần nhận diện, với các kích thước, góc độ, ánh sáng và bối cảnh khác nhau.

- **Mô tả Output của bài toán:** Output của bài toán là tập hợp các bounding box với mỗi box được gắn nhãn tương ứng với một trong 5 đối tượng đã được định trước. Mỗi bounding box sẽ có thông tin về vị trí (tọa độ x, y của góc trên bên trái và góc dưới bên phải) và loại đối tượng mà nó bao quanh.

Ví dụ minh họa: Giả sử trong tập dữ liệu có ảnh chứa một chú mèo. Mô hình sẽ cần phải nhận diện được đối tượng và vẽ bounding box: bao quanh mèo với nhãn tương ứng là "Mèo".

- Giới thiệu Dataset được sử dụng để giải quyết bài toán: Dataset là một tập dữ liệu tùy chỉnh được thu thập qua Google và Kaggle và gắn nhãn cẩn thận để bao gồm đa dạng các trường hợp của 5 đối tượng cần nhận diện. Mỗi ảnh trong dataset cần được gắn nhãn với thông tin về vị trí và loại của các đối tượng xuất hiện trong ảnh, thường được biểu diễn dưới dạng tọa độ của bounding box và nhãn tương ứng.

Ví dụ: Nếu đối tượng là các loại phương tiện (ô tô, xe đạp), động vật (mèo, chó), con người dataset cần chứa các ảnh với sự xuất hiện của những đối tượng này trong các điều kiện ánh sáng, bối cảnh, và góc chụp khác nhau, đi kèm với tập tin annotations mô tả vị trí và loại đối tượng.

Quá trình chuẩn bị dataset đóng vai trò quan trọng trong việc xây dựng và đánh giá mô hình, đảm bảo mô hình có khả năng áp dụng vào thực tế với độ chính xác cao.

* 1. Giới hạn nghiên cứu

Số lượng đối tượng: Phần mềm chỉ có thể nhận dạng và phân loại 5 đối tượng cụ thể là: xe máy, xe đạp, con người, chó và mèo. Các đối tượng khác ngoài phạm vi này sẽ có thể dẫn đến sai lệch.

Độ phức tạp của hình ảnh: Phần mềm có thể gặp khó khăn khi xử lý hình ảnh có độ phức tạp cao, chẳng hạn như hình ảnh có nền phức tạp, đối tượng bị che khuất, hoặc hình ảnh bị nhoè, nền và vật thể cùng màu cũng như các đối tượng có đa dạng kiểu .

Phân giải hình ảnh: Phần mềm có thể gặp khó khăn trong việc xử lý hình ảnh có độ phân giải thấp hoặc cao vượt quá giới hạn được xác định trước.

* 1. Bố cục đồ án

Đồ án được tổ chức thành ba chương chính.

Chương 1: Giới thiệu bài toán nhằm nhận dạng và phân loại năm đối tượng dụng cụ học tập từ các hình ảnh, bao gồm:

1.1 Câu hỏi nghiên cứu, phát biểu rõ ràng về bài toán cần giải quyết và giới thiệu dataset được sử dụng.

1.2 Giới hạn nghiên cứu, đưa ra các ràng buộc và giới hạn của bài toán;

1.3 Bố cục đồ án, trình bày tổng quan về cấu trúc và nội dung của đồ án.

Chương 2: Trình bày giải pháp đề xuất, bao gồm các mục sau:

2.1 Phân tích dữ liệu, mô tả quá trình phân tích và tiền xử lý dữ liệu;

2.2 Mô hình, trình bày cụ thể về kiến trúc của mạng nơ-ron sâu được sử dụng.

Chương 3: Tập trung vào thực nghiệm, mô tả quá trình thử nghiệm và phân tích kết quả thu được.

1. Giải pháp đề xuất
   1. Phân tích dữ liệu

Bộ dữ liệu này có 5005 hình ảnh, kích cỡ 1,1GB về 5 đối tượng lần lượt là:

Bike: 1088 ảnh

Cars: 1143 ảnh

Cats: 925 ảnh

Dogs: 925 ảnh

Human: 924 ảnh

Tập train: 3502 ảnh

Tập test: 502 ảnh

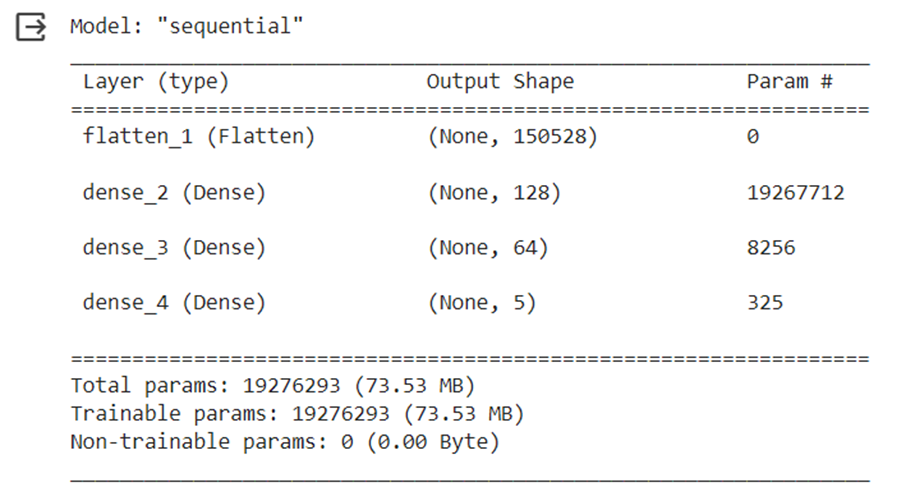
Tập valid: 1001 ảnh

Bộ dữ liệu gồm nhiều hình ảnh với kích thước khác nhau, để tiết kiệm tài nguyên, nhóm em đã giảm độ phân giải của hình ảnh lại.

* 1. Mô hình
     1. Mô hình MLP
* Kiến trúc:



* Mô tả chi tiết kiến trúc:



* Hàm activation của tầng cuối: softmax
* Hàm loss: categorical\_crossentropy
  + 1. Mô hình CNN
* Kiến trúc:

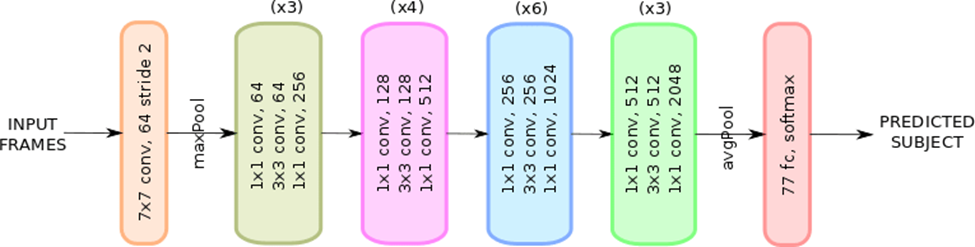


* Mô tả chi tiết kiến trúc:

A screenshot of a computer program

Description automatically generated

* Hàm activation của tầng cuối: softmax.
* Hàm loss: categorical\_crossentropy.
  + 1. Mô hình ResNet 50
* Kiến trúc:



Kiến trúc sâu: ResNet-50 là một mạng nơ-ron sâu với tổng cộng 50 lớp, bao gồm các lớp tích chập, lớp gộp và lớp kết nối đầy đủ.

Khối Residual (Residual Block): Một phần quan trọng của ResNet là các khối residual, mỗi khối này giúp giảm thiểu vấn đề vanishing gradient và giúp việc huấn luyện mạng sâu trở nên dễ dàng hơn. Các khối residual cho phép mạng học được các biểu diễn tốt hơn bằng cách thêm các kết nối dư (residual connections) hoặc skip connections. ResNet-50 sử dụng các khối residual được xếp chồng lên nhau.

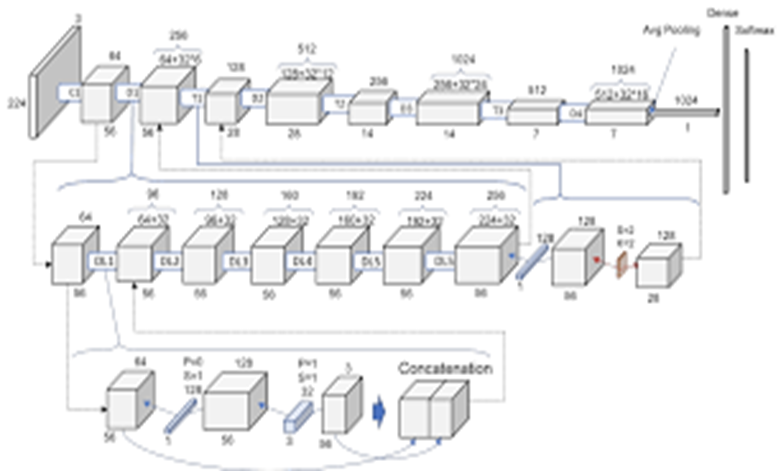
Các kiến trúc chính:

ResNet-50 bao gồm các lớp tích chập và lớp gộp để trích xuất đặc trưng từ dữ liệu đầu vào.

Các khối residual, bao gồm các lớp tích chập, giúp mạng học các biểu diễn sâu hơn.

Cuối cùng, có một hoặc nhiều lớp kết nối đầy đủ để tổng hợp thông tin và thực hiện phân loại.

* + 1. Mô hình DenseNet



Kết nối mật độ cao: DenseNet sử dụng kết nối mật độ cao giữa các lớp, nơi mỗi lớp được kết nối với tất cả các lớp trong trước nó và sau nó. Điều này tạo ra một mạng nơ-ron rất dày đặc và giúp thông tin dễ dàng truyền từ các lớp đến các lớp khác nhau.

Khối Dense: Đặc điểm chính của DenseNet là sự tổ chức của các khối Dense, trong đó mỗi khối có một số lượng lớp tích chập được xếp chồng lên nhau. Mỗi lớp tích chập trong khối Dense nhận đầu vào từ tất cả các lớp trước nó trong khối và chuyển đầu ra của mình cho tất cả các lớp sau nó trong khối.

Kết hợp Batch Normalization và Hàm Kích hoạt: DenseNet thường kết hợp giữa Batch Normalization và hàm kích hoạt (ví dụ: ReLU) sau mỗi lớp tích chập để tăng cường học và giảm vanishing gradient.

Phù hợp cho dữ liệu nhỏ: Do cơ chế kết nối mật độ cao, DenseNet thường được sử dụng trong các tác vụ có lượng dữ liệu huấn luyện nhỏ, vì nó giúp chia sẻ thông tin hiệu quả giữa các lớp.

Hiệu suất cao và tiết kiệm tham số: DenseNet thường cho hiệu suất tốt hơn so với các kiến trúc mạng truyền thống với số lượng tham số ít hơn, do cách mà thông tin được chia sẻ và tái sử dụng hiệu quả trong mạng.

1. Thực nghiệm
2. Mô hình MLP:

* Các Siêu Tham Số:
* Hàm loss: Categorical Cross-Entropy.
* Thuật toán tối ưu hóa: Adam.
* Số lượng epochs: 30.
* Biểu Đồ Loss và Accuracy:

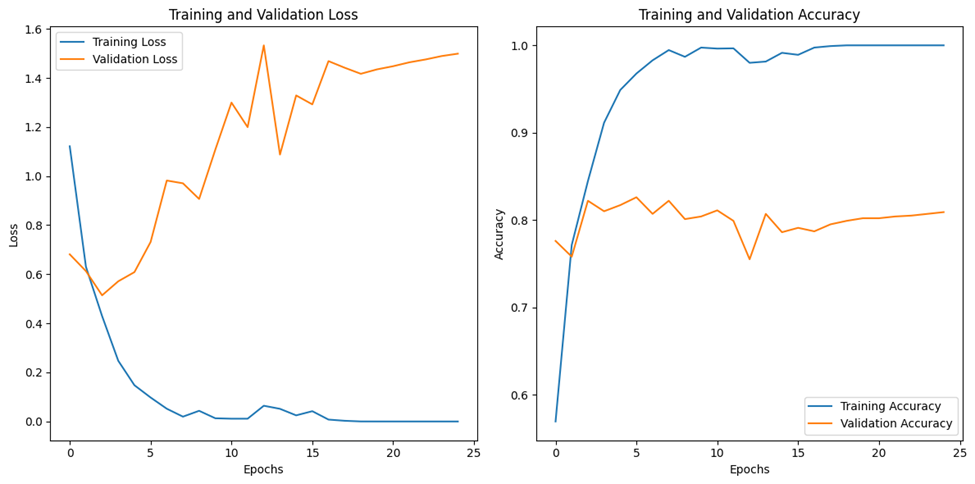


* Kết Quả Thực Nghiệm:

Độ chính xác (Accuracy): 0.428

1. Mô hình CNN:

* Các Siêu Tham Số:
* Hàm loss: Categorical Cross-Entropy.
* Thuật toán tối ưu hóa: Adam.
* Số lượng epochs: 25.
* Biểu Đồ Loss và Accuracy:

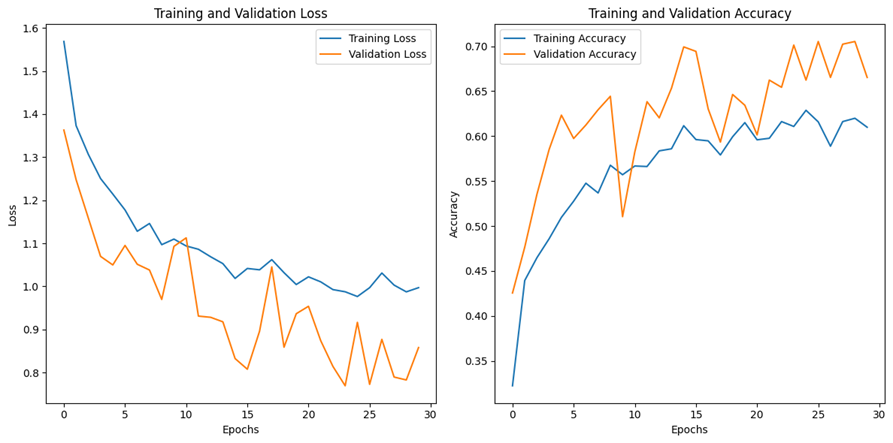


* Kết Quả Thực Nghiệm:

Độ chính xác (Accuracy): 0.745

1. Mô hình Resnet50:

* Các Siêu Tham Số:
* Hàm loss: Categorical Cross-Entropy.
* Thuật toán tối ưu hóa: Adam.
* Số lượng epochs: 30.
* Biểu Đồ Loss và Accuracy:



* Kết Quả Thực Nghiệm:

Độ chính xác (Accuracy): 0.645

1. Mô hình DenseNet:

* Các Siêu Tham Số:
* Hàm loss: Categorical Cross-Entropy.
* Thuật toán tối ưu hóa: Adam.
* Số lượng epochs: 30.
* Biểu Đồ Loss và Accuracy:



* Kết Quả Thực Nghiệm:

Độ chính xác (Accuracy): 0.976

KẾT LUẬN

**Những điều đã làm được**

Trong đồ án này, nhóm đã tiến hành nghiên cứu về việc nhận dạng các đối tượng từ hình ảnh, tập trung vào 5 loại đối tượng chính là bike, cars, cats, dogs, và human. Nhóm đã đề xuất và triển khai một số phương pháp phân loại dựa trên các mô hình MLP, CNN, Resnet50, DenseNet121, và sử dụng một tập dữ liệu chứa 5005 hình ảnh đã được gán nhãn. Sau đó, nhóm đã thực hiện huấn luyện và đánh giá mô hình trên tập dữ liệu kiểm tra để đánh giá hiệu suất của phương pháp đề xuất. Kết quả của nghiên cứu đã cung cấp thông tin quan trọng về khả năng nhận dạng của mô hình trong môi trường thực tế.

**Những điều chưa làm được**

Về phần quay đối tượng về góc đúng đắn chưa áp dụng được nhiều mô hình khác nhau.

**Hướng phát triển**

Có một số hướng phát triển tiềm năng cho đồ án này. Đầu tiên, cần tiến hành một nghiên cứu chi tiết về hiệu suất của mô hình trên từng lớp đối tượng để hiểu rõ hơn về khả năng nhận dạng của mô hình đối với từng loại đối tượng cụ thể. Điều này sẽ giúp xác định các điểm yếu và cải thiện mô hình một cách hiệu quả. Thứ hai, cần tiếp tục đánh giá sự ảnh hưởng của các yếu tố nhiễu như ánh sáng, góc chụp, và nhiễu khác đối với hiệu suất của mô hình để đảm bảo tính ổn định và độ tin cậy trong các điều kiện môi trường khác nhau. Bằng cách tăng cường dữ liệu huấn luyện và tối ưu hóa mô hình, chúng ta có thể cải thiện hiệu suất của mô hình và mở ra những cơ hội mới trong lĩnh vực trí tuệ nhân tạo và tự động hóa.

TÀI LIỆU THAM KHẢO

**Tài liệu Internet**

Great Learning team, 18/11/2022, Introduction to VGG16 | What is VGG16?, <https://www.mygreatlearning.com/blog/introduction-to-vgg16/> , 03/04/2024.

Ứng dụng nhận diện các loài động vật áp dụng thuật toán CNN | 12.04.2022, <https://www.youtube.com/watch?v=MhkZxcsutU0&t=61s>, 03/04/2024.

Thuật toán Convolutional Neural Network, <https://www.youtube.com/watch?v=swguqT77ZLE&list=PL8ZSveYn9kVT0DcOXQKcnuhLlCZeG3a-k>, 03/04/2024