**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

*Trà Vinh, ngày ... tháng ... năm ....*

**Giáo viên hướng dẫn**

*(Ký tên và ghi rõ họ tên)*

**NHẬN XÉT CỦA THÀNH VIÊN HỘI ĐỒNG**

*Trà Vinh, ngày ... tháng ... năm ....*

**Thành viên hội đồng**

*(Ký tên và ghi rõ họ tên)*

# LỜI CẢM ƠN

Trước hết, em xin gửi lời cảm ơn chân thành và sâu sắc nhất đến thầy Nguyễn Mộng Hiền, giảng viên hướng dẫn của em đã tận tình chỉ dạy và hỗ trợ em trong suốt quá trình thực hiện đồ án này. Thầy không chỉ truyền đạt cho em những kiến thức chuyên môn quý giá mà còn chia sẻ những kinh nghiệm thực tiễn, giúp em hiểu sâu hơn về mô hình VGG11 và các kiến thức nền tảng về mạng nơ-ron tích chập. Sự tận tâm và hướng dẫn của thầy đã giúp em tự tin hơn trong từng bước thực hiện, vượt qua những khó khăn và thử thách trong quá trình nghiên cứu.

Em cũng xin gửi lời cảm ơn đến Khoa Kỹ thuật và Công nghệ, Trường Đại học Trà Vinh, đã tạo mọi điều kiện thuận lợi cho em trong quá trình học tập và nghiên cứu. Những kiến thức cơ bản và tài liệu tham khảo mà Khoa cung cấp đã trở thành nền tảng quan trọng, giúp em hiểu rõ và tiếp cận hiệu quả hơn với đồ án này.

Với những kiến thức đã tích lũy được cùng sự hướng dẫn tận tình của thầy, em đã hoàn thành bài báo cáo đồ án cơ sở ngành. Tuy nhiên, do giới hạn về thời gian, kỹ năng và tính phức tạp của đề tài, em nhận thấy vẫn còn nhiều thiếu sót cần hoàn thiện. Em rất mong nhận được những góp ý quý báu từ quý thầy cô để tiếp tục phát triển kỹ năng và kiến thức, hoàn thiện bản thân hơn trong những đồ án tiếp theo.

Em xin chân thành cảm ơn!

# MỤC LỤC

[LỜI CẢM ƠN 3](#_Toc155675161)

[MỤC LỤC 4](#_Toc155675162)

[DANH MỤC HÌNH 6](#_Toc155675163)

[DANH MỤC TỪ VIẾT TẮT 7](#_Toc155675164)

[MỞ ĐẦU 9](#_Toc155675165)

[CHƯƠNG 1: TỔNG QUAN 11](#_Toc155675166)

[CHƯƠNG 2: NGHIÊN CỨU LÝ THUYẾT 12](#_Toc155675167)

[CHƯƠNG 3: HIỆN THỰC HÓA NGHIÊN CỨU 26](#_Toc155675191)

[CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU 32](#_Toc155675197)

[CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 36](#_Toc155675202)

[DANH MỤC TÀI LIỆU THAM KHẢO 38](#_Toc155675206)

[PHỤ LỤC 39](#_Toc155675207)

# DANH MỤC HÌNH

[Đang cập nhật]

# DANH MỤC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **STT** | **KÝ HIỆU VIẾT TẮT** | **NỘI DUNG VIẾT TẮT** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**TÓM TẮT ĐỒ ÁN CƠ SỞ NGÀNH**

Đồ án "Tìm hiểu mô hình VGG11 và ứng dụng" nhằm nghiên cứu sâu về kiến trúc và cơ chế hoạt động của mạng nơ-ron tích chập VGG11 trong nhiệm vụ nhận dạng và phân loại hình ảnh. Đề tài tập trung vào việc phân tích chi tiết cấu trúc các lớp tích chập và kết nối đầy đủ của VGG11, đồng thời triển khai mô hình này trên tập dữ liệu CIFAR-100 với 100 nhãn và 60.000 hình ảnh. Việc sử dụng nền tảng Google Colab cùng các thư viện Python như TensorFlow và PyTorch được áp dụng để tối ưu hóa quá trình huấn luyện và đánh giá mô hình.

Trọng tâm của nghiên cứu là đánh giá hiệu suất của VGG11 trong việc phân loại các đối tượng phức tạp trên tập dữ liệu CIFAR-100. Mục đích chính là hiểu rõ hơn về khả năng của VGG11 trong việc xử lý và phân loại hình ảnh, từ đó xác định những ưu điểm và hạn chế của mô hình trong các ứng dụng thực tiễn.

Kết quả cuối cùng cho thấy mô hình VGG11 đạt được độ chính xác cao trong việc phân loại hình ảnh trên tập dữ liệu CIFAR-100, chứng minh tính hiệu quả của kiến trúc này trong các nhiệm vụ thị giác máy tính. Việc triển khai trên Google Colab đã tối ưu hóa tài nguyên tính toán mà không cần đầu tư vào phần cứng chuyên dụng, tăng tính khả thi và hiệu quả cho quá trình nghiên cứu. Đồ án không chỉ cung cấp cái nhìn sâu sắc về mô hình VGG11 mà còn góp phần tích lũy kiến thức và kinh nghiệm thiết thực cho các ứng dụng AI và thị giác máy tính trong tương lai.

# MỞ ĐẦU

**1. Lý do chọn đề tài**

Trong bối cảnh trí tuệ nhân tạo (AI) và thị giác máy tính ngày càng phát triển mạnh mẽ, việc nghiên cứu và ứng dụng các mô hình học sâu trong nhận dạng và phân loại hình ảnh trở nên vô cùng thiết yếu. Những tiến bộ trong lĩnh vực này đã mang đến nhiều giải pháp thực tiễn, từ công nghệ nhận diện khuôn mặt đến phân loại và xử lý hình ảnh trong y tế và công nghiệp. Các kiến trúc mạng nơ-ron tích chập (CNN) như VGG11 đã chứng minh hiệu quả vượt trội, đặc biệt trong các nhiệm vụ nhận dạng đối tượng với độ chính xác cao, từ đó thúc đẩy việc nghiên cứu sâu hơn về khả năng phân loại ảnh.

Đề tài "Tìm hiểu mô hình VGG11 và ứng dụng" tập trung vào việc nghiên cứu chi tiết cấu trúc và cơ chế hoạt động của VGG11, đồng thời triển khai mô hình này trên tập dữ liệu CIFAR-100. VGG11, với cấu trúc đơn giản nhưng mạnh mẽ, được ứng dụng rộng rãi trong các hệ thống nhận dạng và phân loại ảnh, đáp ứng tốt yêu cầu về độ chính xác cao. Tập dữ liệu CIFAR-100, bao gồm 100 nhãn và 60.000 hình ảnh, cung cấp môi trường thử nghiệm phong phú, giúp kiểm chứng hiệu suất của VGG11 trong việc phân loại các đối tượng phức tạp và đa dạng.

Ngoài ra, việc triển khai mô hình trên nền tảng Google Colab mang lại lợi thế tối ưu hóa tài nguyên tính toán mà không cần đầu tư vào phần cứng chuyên dụng, từ đó nâng cao tính khả thi và hiệu quả cho quá trình nghiên cứu.

**2. Mục tiêu nghiên cứu**

* Tìm hiểu kiến trúc và hoạt động VGG11 trong phân loại hình ảnh.
* Tìm hiểu và tiền xử lý dữ liệu CIFAR-100, đảm bảo phù hợp đầu vào VGG11.
* Triển khai và huấn luyện mô hình VGG11 trên Google Colab.
* Đánh giá hiệu quả của VGG11 qua chỉ số độ chính xác trên CIFAR-100.

**3. Phương pháp nghiên cứu**

* Phân tích cấu trúc và các lớp của VGG11.
* Chuẩn hóa và chuyển đổi dữ liệu CIFAR-100 cho VGG11.
* Triển khai VGG11 trên Google Colab sử dụng ngôn ngữ lập trình Python.
* Đánh giá độ chính xác và hiệu quả của VGG11 trên CIFAR-100.

**4. Đối tượng nghiên cứu**

**Mô hình VGG11:** Kiến trúc mạng nơ-ron tích chập (CNN) với cấu trúc đặc thù, chuyên dùng trong bài toán phân loại hình ảnh.

**Tập dữ liệu CIFAR-100:** Bộ dữ liệu gồm 60.000 hình ảnh thuộc 100 nhãn, được sử dụng để đánh giá khả năng nhận dạng và phân loại của mô hình VGG11.

**Các phương pháp tiền xử lý và huấn luyện mô hình:** Bao gồm các kỹ thuật chuẩn hóa, tối ưu hóa tham số nhằm nâng cao độ chính xác của mô hình VGG11 trên dữ liệu thử nghiệm thông qua Google Colab.

**5. Phạm vi nghiên cứu**

Nghiên cứu tập trung vào cấu trúc và khả năng phân loại của mô hình VGG11 trên tập dữ liệu CIFAR-100 (100 nhãn, 60.000 ảnh), triển khai trên Google Colab với các thư viện Python (TensorFlow hoặc PyTorch), và đánh giá hiệu quả qua chỉ số độ chính xác.

# CHƯƠNG 1: TỔNG QUAN

Trong thời đại phát triển của trí tuệ nhân tạo (AI) và thị giác máy tính, nhu cầu về các hệ thống tự động phân loại và nhận dạng hình ảnh ngày càng trở nên cấp thiết. Các ứng dụng từ nhận diện khuôn mặt, chẩn đoán y khoa cho đến kiểm tra chất lượng sản phẩm trong công nghiệp đều yêu cầu độ chính xác và hiệu suất cao trong việc xử lý và phân loại dữ liệu hình ảnh. Để đáp ứng yêu cầu này, các kiến trúc mạng nơ-ron tích chập (CNN) như VGG11 đã chứng minh tính hiệu quả vượt trội, đặc biệt là trong các nhiệm vụ phân loại đối tượng phức tạp. Đề tài này được chọn nhằm khám phá và nghiên cứu chi tiết mô hình VGG11, một trong những kiến trúc CNN nổi bật và dễ tiếp cận, giúp hiểu rõ hơn về cấu trúc và khả năng ứng dụng trong thị giác máy tính.

Để xây dựng và triển khai mô hình, đề tài sử dụng ngôn ngữ lập trình Python cùng với các thư viện học sâu phổ biến như TensorFlow và PyTorch. Nền tảng Google Colab được lựa chọn làm môi trường thực thi nhờ vào khả năng cung cấp tài nguyên tính toán mạnh mẽ (GPU và TPU) mà không cần đầu tư vào phần cứng chuyên dụng. Tập dữ liệu CIFAR-100, với 100 lớp đối tượng và 60.000 hình ảnh, là cơ sở để huấn luyện và đánh giá hiệu suất của mô hình VGG11 trong việc nhận dạng và phân loại hình ảnh. Quy trình thực hiện bao gồm các bước như tiền xử lý dữ liệu, thiết lập mô hình, huấn luyện và đánh giá, nhằm cung cấp cái nhìn toàn diện về khả năng của VGG11.

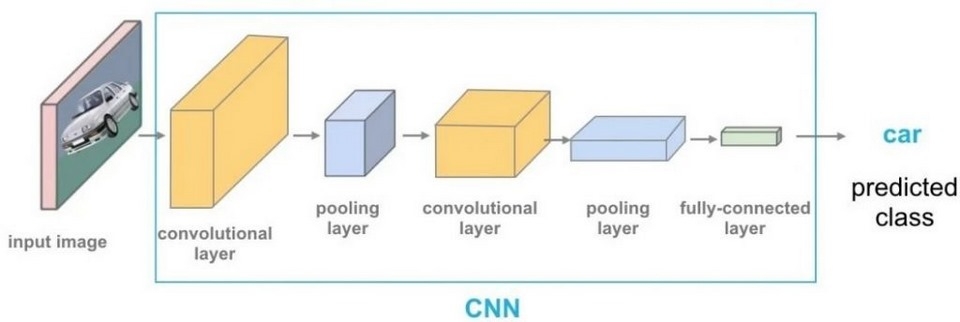
Kết quả sau quá trình nghiên cứu và thử nghiệm cho thấy mô hình VGG11 đạt được độ chính xác cao trên tập dữ liệu CIFAR-100, chứng minh khả năng học sâu và phân loại hiệu quả của kiến trúc này. Mô hình có thể triển khai ở các hệ thống nhận diện hình ảnh trong nhiều lĩnh vực như y tế, giám sát an ninh, và sản xuất công nghiệp, với yêu cầu thời gian chạy hợp lý và độ chính xác cao. Kết quả nghiên cứu không chỉ góp phần vào việc hiểu rõ hơn về VGG11 mà còn tạo nền tảng cho những ứng dụng AI tiềm năng trong tương lai.

# CHƯƠNG 2: NGHIÊN CỨU LÝ THUYẾT

## 2.1. Giới thiệu về mạng Nơ-ron tích chập CNN

### 2.1.1. Khái niệm cơ bản về CNN

Mạng Nơ-ron tích chập (Convolutional Neural Networks - CNN) là một nhánh quan trọng của học sâu (deep learning), được thiết kế đặc biệt để xử lý dữ liệu có cấu trúc dạng lưới, chẳng hạn như hình ảnh. CNN đã chứng minh hiệu quả vượt trội trong nhiều nhiệm vụ thị giác máy tính, bao gồm nhận dạng đối tượng, phân loại hình ảnh, và phát hiện đối tượng. Khả năng tự động học các đặc trưng từ dữ liệu mà không cần sự can thiệp của con người đã làm cho CNN trở thành công cụ ưu việt trong việc giải quyết các vấn đề phức tạp trong lĩnh vực trí tuệ nhân tạo.



Hình 2.1. Cấu trúc cơ bản của CNN

CNN được xây dựng dựa trên mô hình nơ-ron nhân tạo truyền thống, nhưng với các lớp tích chập (convolutional layers) và lớp kết nối toàn cục (fully connected layers) được sắp xếp một cách hệ thống. Các lớp tích chập chịu trách nhiệm trích xuất các đặc trưng không gian từ hình ảnh đầu vào, trong khi các lớp kết nối toàn cục thực hiện việc phân loại dựa trên các đặc trưng đã được trích xuất. Sự kết hợp giữa các lớp này cho phép CNN học các mô hình phức tạp và biểu diễn các mối quan hệ không gian trong dữ liệu hình ảnh một cách hiệu quả.

Ngoài ra, CNN còn sử dụng các kỹ thuật như lớp pooling để giảm kích thước không gian của dữ liệu, giúp giảm số lượng tham số cần học và ngăn ngừa hiện tượng overfitting. Các hàm kích hoạt như ReLU (Rectified Linear Unit) được sử dụng để giới thiệu tính phi tuyến vào mô hình, tăng khả năng biểu diễn của mạng.

### 2.1.2. Nguyên lý hoạt động của CNN

Nguyên lý hoạt động của CNN dựa trên việc mô phỏng cách thức hoạt động của hệ thống thị giác con người, nơi các tế bào thần kinh trong não xử lý thông tin hình ảnh qua nhiều tầng khác nhau. Mỗi tầng trong CNN thực hiện một chức năng cụ thể nhằm trích xuất và học các đặc trưng từ dữ liệu hình ảnh đầu vào.

A diagram of a layer of layers

Description automatically generated

Hình 2.2. Minh họa quá trình hoạt động của lớp tích chập

**Lớp tích chập (Convolutional Layer):** Lớp tích chập là thành phần chính của CNN, chịu trách nhiệm phát hiện các đặc trưng cục bộ trong hình ảnh. Mỗi lớp tích chập bao gồm một số lượng lớn các bộ lọc (filters) nhỏ, thường có kích thước 3x3 hoặc 5x5. Mỗi bộ lọc này di chuyển qua toàn bộ hình ảnh đầu vào theo một bước di chuyển (stride), thực hiện phép nhân ma trận giữa bộ lọc và vùng dữ liệu tương ứng, sau đó tổng hợp kết quả để tạo ra một bản đồ đặc trưng (feature map). Quá trình này giúp CNN phát hiện các đặc trưng như cạnh, góc, và các mẫu lặp lại trong hình ảnh.

**Lớp kích hoạt (Activation Layer):** Sau mỗi lớp tích chập, một hàm kích hoạt phi tuyến như ReLU được áp dụng để giới thiệu tính phi tuyến vào mô hình. Hàm kích hoạt giúp mạng học các mô hình phức tạp hơn bằng cách biến đổi đầu ra của các lớp tích chập, từ đó tăng khả năng biểu diễn của CNN.

**Lớp pooling** **(Pooling Layer):** Lớp pooling được sử dụng để giảm kích thước không gian của các bản đồ đặc trưng, từ đó giảm số lượng tham số và tính toán cần thiết trong mạng. Các kỹ thuật pooling phổ biến bao gồm max pooling và average pooling. Max pooling chọn giá trị lớn nhất trong một vùng nhỏ của bản đồ đặc trưng, trong khi average pooling tính trung bình của các giá trị trong vùng đó. Việc giảm kích thước không gian giúp ngăn ngừa hiện tượng overfitting và tăng tính khả chuyển của mô hình.

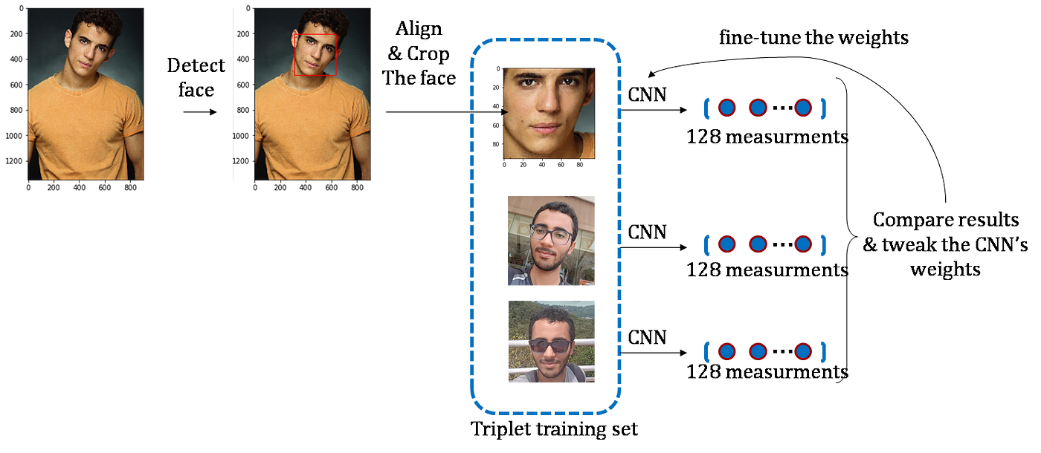
**Lớp kết nối toàn cục (Fully Connected Layer):** Sau các lớp tích chập và các lớp pooling, dữ liệu được đưa vào các lớp kết nối toàn cục, nơi mỗi nơ-ron kết nối với tất cả các nơ-ron trong lớp trước đó. Các lớp này chịu trách nhiệm thực hiện việc phân loại dựa trên các đặc trưng đã được trích xuất từ các lớp tích chập. Kết quả cuối cùng của lớp kết nối toàn cục thường được đưa qua một hàm kích hoạt softmax.

**Quá trình huấn luyện (Training Process):** CNN được huấn luyện thông qua quá trình lan truyền ngược (backpropagation) và tối ưu hóa bằng các thuật toán như Stochastic Gradient Descent (SGD) hoặc Adam. Quá trình huấn luyện nhằm điều chỉnh các trọng số của các bộ lọc và các kết nối trong mạng.

### 2.1.3. Ưu điểm và ứng dụng của CNN trong thị giác máy tính

Mạng Nơ-ron tích chập (CNN) mang lại nhiều ưu điểm vượt trội so với các phương pháp truyền thống trong lĩnh vực thị giác máy tính. Một trong những lợi thế chính của CNN là khả năng tự động học và trích xuất các đặc trưng từ dữ liệu hình ảnh mà không cần sự can thiệp thủ công của con người. Điều này giúp giảm thiểu thời gian và công sức trong việc xử lý dữ liệu, đồng thời tăng độ chính xác và hiệu quả của các mô hình phân loại và nhận dạng.

Một ưu điểm khác của CNN là khả năng xử lý dữ liệu không gian một cách hiệu quả thông qua các lớp tích chập và pooling. Các lớp này không chỉ giúp giảm số lượng tham số cần học mà còn giữ lại các đặc trưng quan trọng của hình ảnh, từ đó cải thiện khả năng tổng quát hóa của mô hình trên các tập dữ liệu mới. Hơn nữa, CNN có khả năng nhận diện các mẫu lặp lại và các mối quan hệ không gian trong hình ảnh, làm cho chúng đặc biệt phù hợp với các nhiệm vụ phân loại và phát hiện đối tượng.



Hình 2.3. Minh họa ứng dụng CNN trong nhận diện khuôn mặt

Trong thực tế, CNN đã được ứng dụng rộng rãi trong nhiều lĩnh vực của thị giác máy tính. Một số ứng dụng tiêu biểu bao gồm:

* **Nhận diện khuôn mặt:** CNN được sử dụng để phát hiện và phân loại khuôn mặt trong các hệ thống an ninh và nhận dạng người dùng trên thiết bị di động.
* **Phân loại hình ảnh y tế:** Trong lĩnh vực y tế, CNN hỗ trợ phân tích hình ảnh y khoa như X-quang, MRI và CT để chẩn đoán bệnh tật một cách chính xác và nhanh chóng.
* **Xử lý ảnh công nghiệp:** CNN giúp tự động hóa các quy trình kiểm tra chất lượng sản phẩm, phát hiện lỗi và phân loại trong dây chuyền sản xuất.
* **Nhận diện phương tiện giao thông:** Ứng dụng trong hệ thống lái xe tự động và quản lý giao thông, giúp phân loại và theo dõi các phương tiện trên đường.
* **Phân loại và tìm kiếm hình ảnh trực tuyến:** Các công cụ tìm kiếm hình ảnh sử dụng CNN để phân loại và gợi ý hình ảnh dựa trên nội dung và đặc trưng của chúng.

Nhờ vào những ưu điểm này, CNN đã trở thành một công cụ mạnh mẽ và linh hoạt trong việc giải quyết các bài toán phức tạp trong thị giác máy tính. Sự phát triển không ngừng của các kiến trúc CNN như VGG11, ResNet và Inception tiếp tục mở ra nhiều cơ hội mới cho việc ứng dụng AI vào các lĩnh vực khác nhau, từ y tế đến công nghiệp và an ninh, góp phần nâng cao hiệu quả và chất lượng cuộc sống.

## 2.2. Kiến trúc mô hình VGG11

### 2.2.1. Lịch sử và sự phát triển của VGG11

Mô hình VGG11 là một trong những kiến trúc mạng nơ-ron tích chập (CNN) tiên tiến, được giới thiệu bởi nhóm nghiên cứu Visual Geometry Group (VGG) tại Đại học Oxford vào năm 2014. VGG11 được phát triển nhằm mục tiêu cải thiện hiệu suất của các mô hình CNN trước đó bằng cách tăng độ sâu của mạng mà vẫn duy trì cấu trúc đơn giản và dễ hiểu. Sự phát triển của VGG11 đánh dấu một bước tiến quan trọng trong việc thiết kế các kiến trúc mạng nơ-ron sâu, góp phần nâng cao khả năng phân loại và nhận dạng hình ảnh trong các ứng dụng thị giác máy tính.

Trước sự ra đời của VGG11, các mô hình CNN như AlexNet và ZFNet đã chứng minh được tiềm năng vượt trội của học sâu trong các nhiệm vụ phân loại hình ảnh trên bộ dữ liệu ImageNet. Tuy nhiên, nhóm VGG nhận thấy rằng việc tăng độ sâu của mạng có thể cải thiện đáng kể hiệu suất mà không cần phải thay đổi cấu trúc cơ bản của các lớp tích chập và kết nối đầy đủ. Do đó, VGG11 được thiết kế với 11 lớp học sâu, bao gồm các lớp tích chập nhỏ (3x3) và các lớp kết nối đầy đủ, tạo nên một kiến trúc mạng có khả năng học các đặc trưng phức tạp từ dữ liệu hình ảnh.

Một trong những điểm nổi bật của VGG11 là việc sử dụng các bộ lọc nhỏ và tăng số lượng lớp tích chập để tăng cường khả năng biểu diễn của mô hình. Điều này giúp VGG11 không chỉ đạt được độ chính xác cao mà còn giảm thiểu hiện tượng mất thông tin trong quá trình truyền tải dữ liệu qua các lớp mạng. Bên cạnh đó, việc duy trì kích thước kernel đồng nhất (3x3) trong toàn bộ mạng giúp đơn giản hóa quá trình thiết kế.

Ngoài ra, VGG11 cũng đã trở thành một cơ sở tham khảo quan trọng cho nhiều nghiên cứu sau này trong lĩnh vực học sâu và thị giác máy tính, ảnh hưởng đến cách thức thiết kế và tối ưu các mô hình mạng nơ-ron hiện đại.

### 2.2.2. Cấu trúc chi tiết của VGG11

Mô hình VGG11 được thiết kế với cấu trúc đơn giản nhưng hiệu quả, bao gồm 11 lớp học sâu (deep layers), trong đó có 8 lớp tích chập (convolutional layers) và 3 lớp kết nối đầy đủ (fully connected layers). Mục tiêu của thiết kế này là để tận dụng độ sâu của mạng nhằm học các đặc trưng phức tạp từ dữ liệu hình ảnh, đồng thời vẫn giữ cho kiến trúc đơn giản và dễ triển khai.

Kiến trúc của VGG11 được xây dựng dựa trên việc sử dụng các bộ lọc nhỏ có kích thước 3x3 cho tất cả các lớp tích chập. Kích thước kernel đồng nhất này không chỉ giúp đơn giản hóa quá trình triển khai mà còn làm tăng khả năng biểu diễn của mạng. Mỗi lớp tích chập sử dụng một số lượng lớn các bộ lọc (tăng dần từ 64 đến 512 bộ lọc qua các tầng mạng), với mục đích học các đặc trưng ở nhiều cấp độ khác nhau, từ các đặc trưng cục bộ (như cạnh và góc) đến các đặc trưng toàn cục phức tạp hơn. Cấu trúc của VGG11 được mô tả chi tiết như sau:

A diagram of a computer generated diagram

Description automatically generated with medium confidence

Hình 2.4. Minh họa cấu trúc chi tiết của VGG11

**Lớp 1 và 2:** Hai lớp tích chập đầu tiên, mỗi lớp gồm 64 bộ lọc kích thước 3x3, giúp trích xuất các đặc trưng cơ bản của hình ảnh đầu vào.

**Lớp 3 và 4:** Hai lớp tích chập kế tiếp với 128 bộ lọc 3x3, giúp tăng khả năng biểu diễn của mạng bằng cách học các đặc trưng ở cấp độ phức tạp hơn. Lớp max pooling tiếp tục được áp dụng sau lớp này để giảm kích thước không gian.

**Lớp 5, 6, 7, và 8:** Đây là bốn lớp tích chập tiếp theo, mỗi lớp bao gồm 256 bộ lọc 3x3. Các lớp này chịu trách nhiệm học các đặc trưng phức tạp hơn nữa, giúp mô hình nhận diện được các mẫu đặc trưng từ dữ liệu đầu vào. Một lớp max pooling được thêm vào sau hai lớp tích chập đầu tiên trong nhóm này để giảm thiểu kích thước.

**Lớp 9 và 10:** Hai lớp tích chập cuối cùng với 512 bộ lọc 3x3. Các lớp này giúp học các đặc trưng toàn cục cao cấp, cần thiết cho việc phân loại các đối tượng phức tạp trong hình ảnh. Một lớp max pooling tiếp tục được áp dụng sau hai lớp tích chập này.

**Lớp kết nối đầy đủ (Fully Connected Layers):** Sau các lớp tích chập và pooling, dữ liệu được đưa qua ba lớp kết nối đầy đủ (fully connected layers), với số lượng nơ-ron giảm dần qua từng lớp. Hai lớp đầu tiên có 4096 nơ-ron, trong khi lớp cuối cùng có số nơ-ron tương ứng với số lớp phân loại trong bài toán cụ thể, thường là 1000 nơ-ron khi áp dụng cho ImageNet.

**Hàm kích hoạt và Softmax:** Mỗi lớp tích chập và lớp kết nối đầy đủ đều sử dụng hàm kích hoạt phi tuyến ReLU, giúp tăng cường khả năng biểu diễn của mô hình. Cuối cùng, lớp kết nối đầy đủ cuối cùng sử dụng hàm softmax để chuyển đổi đầu ra thành xác suất cho mỗi lớp phân loại.

Với cấu trúc đồng nhất và độ sâu vừa phải, VGG11 đạt được hiệu suất phân loại hình ảnh tốt mà không đòi hỏi quá nhiều tham số phức tạp. Cách sử dụng các bộ lọc nhỏ 3x3 và lớp pooling giúp tối ưu hóa tài nguyên tính toán, đồng thời giảm thiểu hiện tượng overfitting trên các tập dữ liệu lớn như ImageNet. Nhờ vào cấu trúc này, VGG11 có khả năng học các đặc trưng hiệu quả, từ đặc trưng cục bộ đến các đặc trưng toàn cục cao cấp, giúp mô hình phù hợp với nhiều bài toán thị giác máy tính khác nhau.

### 2.2.3. So sánh VGG11 với các kiến trúc CNN khác

Mô hình VGG11, với thiết kế đơn giản và độ sâu vừa phải, đã đặt nền tảng cho nhiều kiến trúc CNN hiện đại. Tuy nhiên, so với các mô hình CNN khác như AlexNet, ResNet và Inception, VGG11 có những đặc điểm riêng biệt cả về cấu trúc lẫn hiệu suất. Phần này sẽ so sánh VGG11 với các kiến trúc CNN nổi bật khác để làm rõ điểm mạnh và hạn chế của VGG11.

**So sánh với AlexNet:** AlexNet là một trong những kiến trúc CNN đầu tiên nổi bật, chiến thắng trong cuộc thi ImageNet vào năm 2012 và mở ra kỷ nguyên của học sâu trong thị giác máy tính. AlexNet bao gồm 5 lớp tích chập và 3 lớp kết nối đầy đủ, với kích thước kernel lớn hơn (11x11 và 5x5) so với VGG11. Ngược lại, VGG11 sử dụng các bộ lọc nhỏ hơn (3x3) nhưng tăng số lượng lớp tích chập lên 8 lớp, điều này giúp mô hình có khả năng học các đặc trưng phức tạp hơn.

* **Ưu điểm của VGG11 so với AlexNet:** VGG11 có độ sâu lớn hơn và cấu trúc đồng nhất hơn, giúp mô hình trích xuất đặc trưng chi tiết hơn so với AlexNet.
* **Nhược điểm:** VGG11 yêu cầu tài nguyên tính toán lớn hơn so với AlexNet do số lượng lớp tích chập nhiều hơn và cần nhiều tham số hơn.

**So sánh với ResNet:** ResNet là một cải tiến quan trọng trong thiết kế CNN, nổi bật với khái niệm residual connections (kết nối dư), cho phép truyền thông tin qua các lớp mà không gặp vấn đề gradient vanishing (mất gradient) khi mạng càng sâu. ResNet thường được xây dựng với độ sâu cực lớn, như ResNet50, ResNet101, với các lớp kết nối dư cho phép mạng học hiệu quả hơn khi số lượng lớp tăng lên.

* **Ưu điểm của ResNet so với VGG11:** ResNet có khả năng học hiệu quả ở độ sâu lớn hơn nhờ vào các kết nối dư, giúp mô hình học các đặc trưng phức tạp mà không gặp hiện tượng mất gradient.
* **Nhược điểm của VGG11 so với ResNet:** VGG11 không có các kết nối dư, khiến mô hình gặp khó khăn hơn trong việc học các đặc trưng.

**So sánh với Inception:** Inception (còn gọi là GoogLeNet) là một kiến trúc CNN phức tạp hơn, sử dụng khối Inception để tối ưu hóa hiệu suất tính toán. Thay vì chỉ sử dụng các lớp tích chập đồng nhất, Inception kết hợp các bộ lọc với kích thước khác nhau trong cùng một lớp, giúp mô hình học nhiều đặc trưng khác nhau đồng thời, từ đó tăng khả năng biểu diễn của mạng mà vẫn giữ được số lượng tham số hợp lý.

* **Ưu điểm của Inception so với VGG11:** Inception có khả năng xử lý hiệu quả các đặc trưng đa dạng nhờ khối Inception, trong khi vẫn tiết kiệm tài nguyên tính toán nhờ vào thiết kế tối ưu.
* **Nhược điểm của VGG11 so với Inception:** VGG11 có cấu trúc đơn giản và đồng nhất, điều này dễ triển khai nhưng kém linh hoạt hơn trong việc xử lý nhiều loại đặc trưng khác nhau.

### 2.2.4. Ưu điểm và nhược điểm của VGG11

**Ưu điểm nổi bật của VGG11:** VGG11 được đánh giá cao nhờ vào cấu trúc đơn giản nhưng mạnh mẽ, với các bộ lọc đồng nhất 3x3 và thiết kế lớp tích chập sâu. Cấu trúc này giúp VGG11 dễ triển khai, mở rộng và hiểu rõ hơn so với các mô hình phức tạp khác, đồng thời vẫn đạt được hiệu suất cao trên các bài toán phân loại.

**Nhược điểm của VGG11:** Mặc dù VGG11 có khả năng biểu diễn mạnh mẽ, nhưng nó đòi hỏi tài nguyên tính toán lớn do số lượng tham số cao và thiếu các cơ chế tối ưu hóa như kết nối dư của ResNet hoặc khối Inception. Điều này khiến VGG11 không phù hợp cho các ứng dụng có giới hạn về tài nguyên tính toán.

## 2.3. Tiêu chí đánh giá hiệu suất mô hình VGG11

### 2.3.1. Các chỉ số đánh giá hiệu suất trong phân loại hình ảnh

Đánh giá hiệu suất của mô hình phân loại hình ảnh thường dựa vào một số chỉ số chính nhằm đo lường độ chính xác và khả năng tổng quát hóa của mô hình. Các chỉ số phổ biến bao gồm:

* **Độ chính xác (Accuracy):** Đo lường tỷ lệ phần trăm các dự đoán đúng trên tổng số dự đoán. Đây là chỉ số cơ bản nhất để đánh giá hiệu suất của mô hình phân loại.
* **Độ chính xác từng lớp (Per-Class Accuracy):** Đo lường độ chính xác của mô hình đối với từng lớp cụ thể, giúp xác định các lớp mà mô hình hoạt động tốt hoặc gặp khó khăn.
* **Precision, Recall và F1-Score:** Các chỉ số này cung cấp cái nhìn sâu sắc hơn về hiệu suất của mô hình, đặc biệt trong các tình huống dữ liệu không cân bằng. Precision đo lường tỷ lệ các dự đoán đúng trong số các dự đoán dương, Recall đo lường tỷ lệ các dự đoán dương đúng trong tổng số các trường hợp dương thực sự, và F1-Score là trung bình điều hòa của Precision và Recall.
* **Confusion Matrix (Ma trận nhầm lẫn):** Cung cấp một bảng tổng hợp các dự đoán đúng và sai của mô hình, giúp phân tích chi tiết hơn về cách mà mô hình phân loại các lớp khác nhau.

Việc sử dụng các chỉ số này giúp đánh giá toàn diện hiệu suất của mô hình VGG11, từ đó đưa ra các cải tiến cần thiết để nâng cao độ chính xác và khả năng tổng quát hóa của mô hình.

### 2.3.2. Phương pháp phân tích kết quả huấn luyện

Sau khi huấn luyện mô hình VGG11 trên bộ dữ liệu CIFAR-100, việc phân tích kết quả huấn luyện là bước quan trọng để hiểu rõ hơn về hiệu suất và hành vi của mô hình. Các phương pháp phân tích kết quả bao gồm:

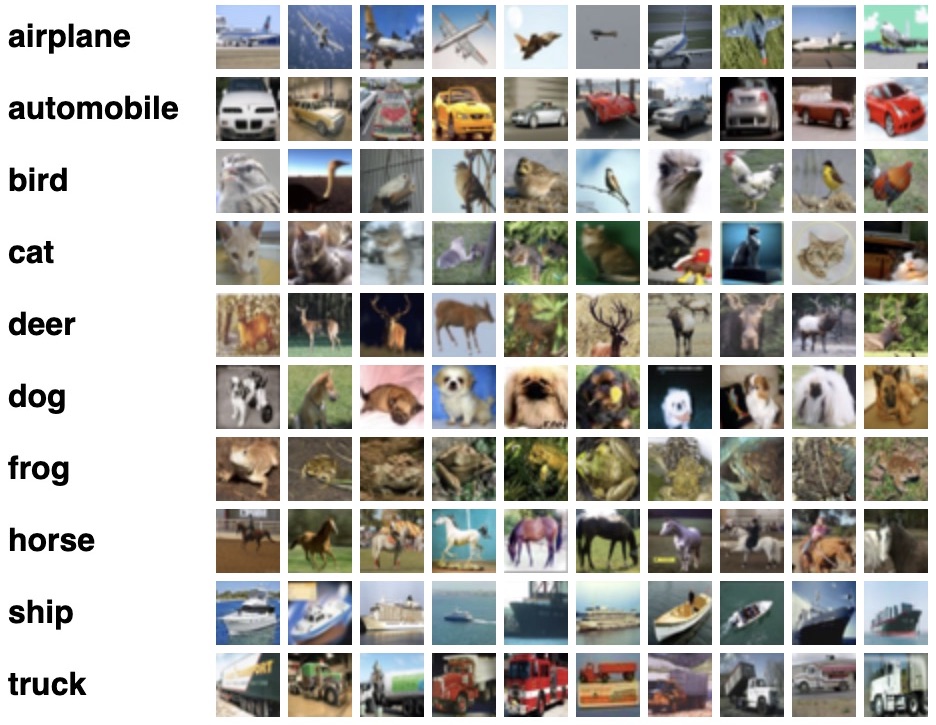
* **Biểu đồ Loss và Accuracy theo Epoch:** Vẽ đồ thị thể hiện sự thay đổi của giá trị loss và accuracy qua từng epoch giúp theo dõi quá trình hội tụ của mô hình. Biểu đồ này giúp xác định khi nào mô hình đạt được điểm dừng huấn luyện hợp lý và liệu mô hình có đang gặp hiện tượng overfitting hay không.
* **Phân tích Ma trận nhầm lẫn (Confusion Matrix):** Sử dụng ma trận nhầm lẫn để phân tích chi tiết các lỗi của mô hình, xác định các lớp mà mô hình thường nhầm lẫn và tìm hiểu nguyên nhân gây ra những nhầm lẫn đó.
* **Visualization của Feature Maps:** Hiển thị các bản đồ đặc trưng (feature maps) từ các lớp khác nhau của mô hình để hiểu rõ hơn về cách mà mô hình học và trích xuất các đặc trưng từ dữ liệu hình ảnh.
* **Đánh giá tổng quát và sâu sắc về các chỉ số:** So sánh các chỉ số đánh giá như Accuracy, Precision, Recall và F1-Score để có cái nhìn toàn diện về hiệu suất của mô hình trên các lớp khác nhau và tổng thể bộ dữ liệu.

Thông qua các phương pháp phân tích này, nhà nghiên cứu có thể đánh giá chính xác hơn về hiệu suất của mô hình VGG11, từ đó đưa ra các chiến lược cải tiến và tối ưu hóa mô hình nhằm nâng cao khả năng phân loại hình ảnh một cách hiệu quả hơn.

## 2.4. Giới thiệu bộ dữ liệu CIFAR-100

### 2.4.1. Mô tả và đặc điểm của CIFAR-100

Bộ dữ liệu CIFAR-100 là một trong những bộ dữ liệu phổ biến và được sử dụng rộng rãi trong lĩnh vực thị giác máy tính, đặc biệt là trong các bài toán phân loại hình ảnh. CIFAR-100 được thiết kế để cung cấp một thách thức đáng kể cho các mô hình học sâu nhờ vào sự đa dạng và phức tạp của các lớp đối tượng. Bộ dữ liệu này không chỉ giúp đánh giá hiệu suất của các mô hình CNN mà còn thúc đẩy sự phát triển của các kỹ thuật tiên tiến trong việc xử lý và phân loại hình ảnh.



Hình 2.5. Minh họa bộ dữ liệu CIFAR-100

Bộ dữ liệu CIFAR-100 bao gồm tổng cộng 60.000 hình ảnh màu, mỗi hình có kích thước 32x32 pixel. Những hình ảnh này được chia thành 100 lớp đối tượng khác nhau, mỗi lớp chứa 600 hình ảnh, trong đó 500 hình dành cho tập huấn luyện và 100 hình dành cho tập kiểm tra. Các lớp đối tượng trong CIFAR-100 được phân thành 20 nhóm siêu lớp (superclasses), mỗi siêu lớp bao gồm 5 lớp đối tượng con, tạo nên sự phân loại đa cấp độ và phức tạp.

Bộ dữ liệu CIFAR-100 còn được sử dụng rộng rãi trong các nghiên cứu và thi đấu về học sâu, cung cấp một nền tảng vững chắc để so sánh hiệu suất của các kiến trúc mạng nơ-ron khác nhau. Sự đa dạng và phức tạp của dữ liệu giúp mô hình CNN phát triển khả năng học các đặc trưng phong phú và cải thiện hiệu suất phân loại trên các tập dữ liệu thực tế.

### 2.4.2. Các phương pháp tiền xử lý dữ liệu cho CIFAR-100

Tiền xử lý dữ liệu là một bước quan trọng trong quá trình xây dựng và huấn luyện mô hình CNN, ảnh hưởng trực tiếp đến hiệu suất và khả năng tổng quát hóa của mô hình. Dưới đây là các phương pháp tiền xử lý dữ liệu phổ biến được áp dụng cho bộ dữ liệu CIFAR-100:

**Chuẩn hóa dữ liệu (Normalization):** Chuẩn hóa dữ liệu là quá trình điều chỉnh các giá trị pixel của hình ảnh để chúng nằm trong một khoảng nhất định, thường là từ 0 đến 1 hoặc có trung bình bằng 0 và độ lệch chuẩn bằng 1. Điều này giúp giảm thiểu sự khác biệt về độ sáng và màu sắc giữa các hình ảnh, tạo điều kiện cho mô hình học sâu tiếp cận dữ liệu một cách hiệu quả hơn.

**Tăng cường dữ liệu (Data Augmentation):** Tăng cường dữ liệu là kỹ thuật mở rộng bộ dữ liệu huấn luyện bằng cách tạo ra các biến thể mới của hình ảnh gốc thông qua các phép biến đổi như xoay, lật, cắt, phóng to, thu nhỏ, và thay đổi độ sáng. Điều này giúp mô hình học được các đặc trưng đa dạng hơn và cải thiện khả năng tổng quát hóa trên các dữ liệu chưa từng thấy.

**Chuyển đổi định dạng (Format Conversion):** Chuyển đổi định dạng hình ảnh từ định dạng PIL sang tensor là bước cần thiết để đưa dữ liệu vào mô hình CNN. Tensor là định dạng dữ liệu phù hợp cho các thư viện học sâu như TensorFlow và PyTorch, cho phép thực hiện các phép toán ma trận hiệu quả.

**Bỏ qua thông tin (Data Cleaning):** Bỏ qua các hình ảnh bị lỗi hoặc không phù hợp là bước quan trọng để đảm bảo chất lượng dữ liệu huấn luyện. Việc loại bỏ các hình ảnh không rõ ràng, mờ nhòe hoặc bị nhiễu giúp mô hình học được các đặc trưng chính xác hơn từ dữ liệu.

**Chia dữ liệu (Data Splitting):** Chia dữ liệu thành các tập huấn luyện, kiểm tra và xác nhận là bước cần thiết để đánh giá hiệu suất của mô hình một cách khách quan. Đảm bảo rằng các tập này được chia đều và không có sự trùng lặp giúp mô hình được đánh giá chính xác về khả năng tổng quát hóa.

Các phương pháp tiền xử lý dữ liệu này không chỉ giúp cải thiện chất lượng dữ liệu mà còn tăng cường khả năng học của mô hình CNN, từ đó nâng cao hiệu suất phân loại trên bộ dữ liệu CIFAR-100. Việc áp dụng các kỹ thuật này một cách hợp lý sẽ giúp mô hình VGG11 khai thác tối đa thông tin từ dữ liệu và đạt được kết quả tốt nhất trong nhiệm vụ phân loại hình ảnh.

## 2.5. Công cụ và thư viện hỗ trợ

### 2.5.1. Python và các thư viện

Python là ngôn ngữ lập trình phổ biến nhất trong lĩnh vực học sâu và trí tuệ nhân tạo nhờ vào cú pháp đơn giản, dễ học và khả năng mở rộng cao. Sự phong phú của các thư viện và công cụ hỗ trợ làm cho Python trở thành lựa chọn ưu việt cho việc phát triển các mô hình mạng nơ-ron tích chập (CNN) như VGG11. Trong đó, TensorFlow và PyTorch là hai thư viện học sâu hàng đầu, được sử dụng rộng rãi trong cộng đồng nghiên cứu và công nghiệp.

TensorFlow do Google phát triển, nổi bật với khả năng triển khai mô hình trên nhiều nền tảng khác nhau, từ máy tính cá nhân đến các nền tảng hệ thống máy chủ lớn. TensorFlow cung cấp một loạt các công cụ mạnh mẽ như TensorBoard để trực quan hóa quá trình huấn luyện và Keras, một API cao cấp giúp đơn giản hóa việc xây dựng và huấn luyện mô hình. Khả năng tối ưu hóa hiệu suất và hỗ trợ mạnh mẽ cho GPU/TPU cũng làm tăng hiệu quả huấn luyện các mô hình lớn như VGG11.

PyTorch, được phát triển bởi Facebook, nổi tiếng với tính linh hoạt và dễ dàng trong việc triển khai các mô hình phức tạp. PyTorch còn hỗ trợ tính năng tính toán động (dynamic computation graph), cho phép thay đổi cấu trúc mạng trong quá trình huấn luyện, điều này rất hữu ích cho việc thử nghiệm và phát triển các kiến trúc mạng mới.

Việc lựa chọn giữa TensorFlow và PyTorch phụ thuộc vào yêu cầu cụ thể của dự án và sở thích cá nhân của nhà nghiên cứu. Cả hai thư viện đều cung cấp đầy đủ các tính năng cần thiết để xây dựng, huấn luyện và đánh giá mô hình VGG11 một cách hiệu quả.

### 2.5.2. Nền tảng Google Colab

Google Colab là một nền tảng đám mây miễn phí, cung cấp môi trường lập trình Python trực tuyến với hỗ trợ mạnh mẽ cho các thư viện học sâu như TensorFlow và PyTorch. Google Colab cho phép người dùng tận dụng sức mạnh của GPU và TPU mà không cần đầu tư vào phần cứng chuyên dụng, giúp giảm chi phí và tăng tốc độ huấn luyện mô hình.



Hình 2.6. Minh họa biểu tượng của Google Colab

Một trong những ưu điểm nổi bật của Google Colab là khả năng chia sẻ và hợp tác dễ dàng. Người dùng có thể chia sẻ notebook với đồng nghiệp hoặc cộng đồng, cho phép cộng tác và trao đổi ý tưởng một cách thuận tiện.

Google Colab cũng cung cấp nhiều tiện ích hỗ trợ khác như khả năng cài đặt các thư viện bổ sung một cách dễ dàng thông qua các lệnh shell, khả năng sử dụng các plugin mở rộng để tăng cường chức năng của notebook, và hỗ trợ tương tác với các dịch vụ đám mây khác. Điều này giúp nhà nghiên cứu linh hoạt trong việc triển khai và thử nghiệm các mô hình học sâu mà không gặp phải các rào cản về mặt kỹ thuật.

Ngoài ra, Google Colab còn hỗ trợ các tính năng như phiên bản lưu tự động, lịch sử chỉnh sửa, và khả năng chạy mã song song, giúp quá trình phát triển và kiểm thử mô hình trở nên suôn sẻ và hiệu quả hơn.

# CHƯƠNG 3: HIỆN THỰC HÓA NGHIÊN CỨU

## 3.1. [Tên]

[Đang cập nhật]

# CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU

## 4.1. [Tên]

[Đang cập nhật]

# CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## 5.1. Kết luận

[Đang cập nhật]

## 5.2. Hạn chế

[Đang cập nhật]

## 5.3. Phương hướng phát triển

[Đang cập nhật]

# DANH MỤC TÀI LIỆU THAM KHẢO

[Đang cập nhật]

# PHỤ LỤC

[Đang cập nhật]