**TRƯỜNG ĐẠI HỌC HÀNG HẢI VIỆT NAM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**-----\*\*\*-----**



**BÁO CÁO BÀI TẬP LỚN**

**HỌC PHẦN “NHẬN DIỆN VÀ XỬ LÝ ẢNH”**

***Đề tài:***

***Xây dựng chương trình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN***

GVHD: TS. Nguyễn Hữu Tuân

Sinh viên thực hiện: Vũ Trung Kiên – 86313

***Hải Phòng, ngày 10 tháng 4 năm 2023***

**MỞ ĐẦU**

Tương lai của nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN là rất tiềm năng và đầy hứa hẹn. Với sự phát triển của công nghệ và trí tuệ nhân tạo, chúng ta có thể nâng cao hiệu suất và độ chính xác của các mô hình nhận dạng để phù hợp với các ứng dụng thực tế khác nhau.

Một trong những ứng dụng tiềm năng của nhận dạng biển báo giao thông là trong lĩnh vực tự động hóa xe hơi, cụ thể là xe tự lái. Nhận dạng biển báo giao thông có thể giúp xe tự động phân biệt và hiểu được các biển báo giao thông trên đường, giúp đảm bảo an toàn cho người lái và người đi đường.

Không chỉ thế, ứng dụng của nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN còn rất nhiều trong thực tiễn. Một số ứng dụng như: hệ thống giám sát giao thông, hệ thống hỗ trợ lái xe tự động, hệ thống định vị vị trí và tìm đường đi dựa trên biển báo giao thông, hệ thống cảnh báo chạm va trước khi xảy ra tai nạn. Tất cả những ứng dụng này đều cần đến việc nhận dạng chính xác các biển báo giao thôngViệc sử dụng mô hình nhận dạng biển báo giao thông sẽ giúp cải thiện khả năng phát hiện các hành vi vi phạm luật giao thông và đảm bảo an toàn cho người tham gia giao thông.

Với báo cáo này, em hy vọng có thể giới thiệu về một trong những ứng dụng thực tế của trí tuệ nhân tạo và đóng góp vào việc nâng cao nhận thức về tính ứng dụng và tiềm năng của công nghệ trong cuộc sống hàng ngày. Ngoài ra, em cũng mong muốn có thể chia sẻ kiến thức về mạng nơ-ron tích chập CNN và sử dụng thư viện Keras và TensorFlow để xây dựng mô hình nhận dạng biển báo giao thông, giúp mọi người có thể tiếp cận với công nghệ này và sử dụng nó trong các ứng dụng của riêng mình.

**MỤC LỤC**

**CHƯƠNG 1**

**GIỚI THIỆU VỀ CHƯƠNG TRÌNH NHẬN DIỆN BIỂN BÁO GIAO THÔNG BẰNG CNN (Convolutional Neural Network)**

1. **Đặt vấn đề**

Trong thời đại 4.0, vấn đề an toàn giao thông trở nên ngày càng quan trọng và đòi hỏi sự chú ý cao đối với việc sử dụng đường bộ. Một trong những yếu tố quan trọng để đảm bảo an toàn giao thông là việc đọc và hiểu biển báo giao thông. Tuy nhiên, do số lượng biển báo giao thông rất lớn và đa dạng, việc nhận diện và hiểu biển báo giao thông đôi khi gặp nhiều khó khăn. Vì vậy, sử dụng nhận dạng biển báo giao thông là một giải pháp hiệu quả giúp cho việc nhận diện và hiểu biển báo giao thông trở nên dễ dàng hơn.

Công dụng của nhận dạng biển báo giao thông là giúp cho việc điều khiển giao thông trở nên dễ dàng và an toàn hơn. Nó có thể được sử dụng trong các hệ thống hỗ trợ lái xe tự động, các ứng dụng quản lý giao thông và các hệ thống an ninh. Các công nghệ nhận dạng biển báo giao thông đang được sử dụng rộng rãi trong các thiết bị di động và các hệ thống nhúng.

Nhận dạng biển báo giao thông đóng vai trò quan trọng trong xã hội hiện nay. Nó có thể giúp giảm thiểu các tai nạn giao thông, tăng cường an toàn đường bộ và giúp người lái xe dễ dàng điều khiển phương tiện.

Chương trình xây dựng nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN là một trong những giải pháp hiệu quả để nhận diện và hiểu biển báo giao thông. Chương trình này được xây dựng trên thư viện TensorFlow và Keras, cho phép tạo ra một mô hình học máy đáng tin cậy để nhận dạng biển báo giao thông. Các bước xây dựng mô hình bao gồm thu thập và tiền xử lý dữ liệu, xây dựng mô hình CNN, huấn luyện và đánh giá mô hình, và sử dụng mô hình để nhận dạng biển báo giao thông trong thực tế.

Với những ưu điểm của mình, chương trình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN sử dụng được rộng rãi trong nhiều lĩnh vực khác nhau như: giúp cải thiện an toàn giao thông, hỗ trợ đối tượng người khuyết tật, hỗ trợ hệ thống giám sát an ninh, phát hiện bất thường trong sản xuất công nghiệp, giúp cải thiện chất lượng cuộc sống và tăng hiệu quả kinh doanh.

Trong lĩnh vực giao thông, chương trình nhận dạng biển báo giao thông giúp giảm thiểu tai nạn giao thông bằng cách nhận diện và cảnh báo tài xế về biển báo giao thông. Điều này đặc biệt hữu ích trong các trường hợp tài xế chưa quen với địa phương hoặc không biết các biển báo đường. Chương trình cũng giúp cải thiện tính an toàn của những người tham gia giao thông, đặc biệt là trong các khu vực đông dân cư hoặc trên đường cao tốc.

Ngoài ra, chương trình còn hỗ trợ đối tượng người khuyết tật, giúp họ có thể tham gia giao thông một cách an toàn hơn. Các hệ thống giám sát an ninh cũng sử dụng chương trình này để phát hiện bất thường hoặc vi phạm trên đường phố.

Với sự phát triển của công nghệ và cuộc sống hiện đại, chương trình nhận dạng biển báo giao thông đóng vai trò rất quan trọng trong việc cải thiện chất lượng cuộc sống và tăng hiệu quả kinh doanh. Ví dụ, các công ty vận tải và các dịch vụ giao hàng đều sử dụng chương trình này để quản lý xe của mình một cách an toàn và hiệu quả hơn.

Tóm lại, chương trình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN đóng vai trò rất quan trọng trong việc cải thiện an toàn giao thông, hỗ trợ đối tượng người khuyết tật, giám sát an ninh và tăng hiệu quả kinh doanh. Chương trình này đã và đang được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau và sẽ tiếp tục phát triển trong tương lai.

1. **Mục tiêu và nhiệm vụ**

Mục tiêu của đề tài xây dựng chương trình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN bao gồm:

* Nghiên cứu và phân tích các kỹ thuật nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN, bao gồm việc hiểu rõ cơ chế hoạt động của mạng nơ-ron tích chập và các phương pháp tiền xử lý dữ liệu cần thiết để đưa dữ liệu vào mạng nơ-ron tích chập.
* Xây dựng tập dữ liệu biển báo giao thông đủ lớn, đa dạng và chất lượng cao để huấn luyện và kiểm tra mô hình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập.
* Thiết kế và huấn luyện mô hình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập, với mục tiêu đạt được độ chính xác cao nhất có thể.
* Kiểm tra và đánh giá hiệu suất của mô hình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập trên tập dữ liệu kiểm tra độc lập và so sánh kết quả với các phương pháp nhận dạng biển báo khác.

Các nhiệm vụ cụ thể của đề tài bao gồm:

* + Nghiên cứu, tìm hiểu về các phương pháp nhận dạng biển báo giao thông.
  + Xây dựng cơ sở dữ liệu cho chương trình nhận dạng.
  + Thiết kế và xây dựng mô hình mạng nơ-ron tích chập CNN cho việc nhận dạng biển báo giao thông.
  + Tiến hành huấn luyện và kiểm thử mô hình trên dữ liệu thực tế.
  + Đánh giá hiệu suất của mô hình nhận dạng được xây dựng.
  + Triển khai chương trình nhận dạng trên nền tảng ứng dụng.

**3. Phương tiện và công cụ để triển khai**

Để triển khai nghiên cứu xây dựng chương trình nhận dạng biển báo giao thông sử dụng mạng nơ-ron tích chập CNN, có thể sử dụng các công cụ và phương tiện sau:

- Ngôn ngữ lập trình Python và các thư viện hỗ trợ như Keras, TensorFlow, PyTorch: Các công cụ này được sử dụng để xây dựng và huấn luyện mô hình mạng nơ-ron tích chập CNN, đưa ra các dự đoán về biển báo giao thông.

- Các tập dữ liệu về biển báo giao thông: Để huấn luyện và kiểm tra mô hình, cần phải sử dụng các tập dữ liệu chứa ảnh của các biển báo giao thông.

- Các công cụ và phần mềm để tiền xử lý dữ liệu ảnh: Các công cụ này được sử dụng để tiền xử lý dữ liệu ảnh trước khi đưa vào huấn luyện mô hình, như tách ảnh ra khỏi các phông nền, thay đổi kích thước ảnh, chuyển đổi định dạng ảnh,...

- Các công cụ và phần mềm để đánh giá mô hình: Để đánh giá độ chính xác của mô hình, có thể sử dụng các công cụ và phần mềm để tính toán độ chính xác, độ nhạy và độ phân giải của mô hình.

- Phần cứng máy tính có cấu hình tương đối cao: Cần sử dụng máy tính có cấu hình tương đối cao để đảm bảo quá trình huấn luyện và kiểm tra mô hình diễn ra nhanh chóng và hiệu quả.

**4. Phương pháp sử dụng phương tiện và công cụ**

Các ứng dụng của mạng nơ-ron tích chập (Convolutional Neural Network - CNN) tập trung vào xử lý ảnh và nhận dạng hình ảnh, được thiết kế để học cách phân tích và trích xuất đặc trưng từ hình ảnh đầu vào bằng cách áp dụng các bộ lọc tích chập và lớp kích hoạt phi tuyến tính. CNN đã được áp dụng rộng rãi trong các ứng dụng như nhận dạng khuôn mặt, phân loại đối tượng, nhận dạng biển số xe, phát hiện và phân tích bệnh trong hình ảnh y tế. Ngoài ra, CNN cũng có thể được sử dụng để xử lý dữ liệu tín hiệu, nhận dạng âm thanh, phân tích dữ liệu tài chính và kinh doanh, phát hiện gian lận, v.v.

TensorFlow là một thư viện mã nguồn mở được phát triển bởi Google Brain Team để xây dựng và huấn luyện các mô hình machine learning và deep learning. TensorFlow cung cấp các API và công cụ để xây dựng các mô hình machine learning và deep learning, bao gồm cả mạng nơ-ron tích chập, mạng nơ-ron hồi quy và các mô hình tạo đối sánh. Nó là một trong những thư viện phổ biến nhất cho việc xây dựng các mô hình deep learning và có nhiều ứng dụng như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, dự báo tài chính, điều khiển robot, v.v. TensorFlow cũng cho phép tính toán song song trên nhiều máy tính và tạo ra các đồ thị luồng dữ liệu để dựng nên các mô hình.

Keras là một thư viện mã nguồn mở cho deep learning được viết bằng Python. Keras cung cấp một giao diện dễ sử dụng để xây dựng và huấn luyện các mô hình deep learning như mạng nơ-ron truyền thẳng, mạng nơ-ron tích chập và mạng nơ-ron hồi quy. Keras được tích hợp sẵn với các thư viện deep learning khác như TensorFlow và Theano để cung cấp các tính năng đầy đủ cho các mô hình deep learning. Nó được sử dụng rộng rãi trong các ứng dụng deep learning và là một trong những thư viện deep learning phổ biến nhất và được sử dụng nhiều nhất trên toàn cầu. Keras được thiết kế để tối giản hóa quá trình xây dựng và huấn luyện mô hình deep learning bằng cách cung cấp các API trừu tượng và dễ hiểu cho người dùng. Với Keras, người dùng có thể dễ dàng xây dựng và tùy chỉnh các lớp mạng nơ-ron, tối ưu hóa và hàm mất mát, và lập lịch huấn luyện. Ngoài ra, Keras còn hỗ trợ các tính năng như việc lưu và tải mô hình, trực quan hóa quá trình huấn luyện, và đánh giá hiệu suất của mô hình. Với tính linh hoạt và dễ sử dụng của mình, Keras là một công cụ hữu ích cho các nhà phát triển và nhà nghiên cứu deep learning, đặc biệt là đối với những người mới bắt đầu.

PyTorch là một thư viện học máy mã nguồn mở được phát triển bởi Facebook, cung cấp cho người dùng các công cụ để xây dựng các mô hình học máy và học sâu. PyTorch được thiết kế để cung cấp một trải nghiệm lập trình tốt hơn cho các nhà phát triển và các nhà nghiên cứu, và cho phép họ dễ dàng xây dựng các mô hình học máy phức tạp bằng Python. PyTorch cung cấp các tính năng và công cụ để xây dựng và huấn luyện các mô hình mạng nơ-ron tích chập và các mô hình học sâu khác, bao gồm cả mạng nơ-ron truyền thẳng và mạng nơ-ron hồi quy. PyTorch cũng cung cấp các tính năng để phân tích và trực quan hóa dữ liệu và kết quả mô hình, giúp người dùng dễ dàng hiểu và tối ưu hóa các mô hình của họ. PyTorch được đánh giá là một trong những thư viện học máy và học sâu được ưa chuộng nhất hiện nay, được sử dụng rộng rãi trong cộng đồng nghiên cứu và công nghiệp.

**CHƯƠNG 2**

**TRÌNH BÀY VÀ PHÂN TÍCH THUẬT TOÁN CNN**

1. **Tổng quan**

Mạng nơ-ron tích chập (CNN) là một loại mạng nơ-ron thường được sử dụng trong học sâu để xử lý dữ liệu liên quan đến hình ảnh và video. Mô hình này được thiết kế để học các đặc trưng cục bộ của ảnh thông qua việc sử dụng các bộ lọc tích chập, và sau đó kết hợp các đặc trưng này để dự đoán nhãn của ảnh đầu vào.

Một mô hình CNN bao gồm một hoặc nhiều lớp tích chập (convolutional layers), các lớp kích hoạt phi tuyến (non-linear activation layers) và các lớp tổng hợp (pooling layers). Các lớp tích chập giúp học các đặc trưng cục bộ của ảnh, trong khi các lớp kích hoạt phi tuyến giúp mô hình học được các đặc trưng phi tuyến tính và các lớp tổng hợp giúp giảm thiểu số lượng thông tin cần xử lý của mô hình.

Sau khi đã học được các đặc trưng cục bộ của ảnh, mô hình CNN sẽ chuyển sang các lớp đầy đủ kết nối (fully-connected layers) để dự đoán nhãn của ảnh. Các lớp đầy đủ kết nối có thể được sử dụng để kết hợp các đặc trưng đã học từ các lớp tích chập để tạo ra một bộ đặc trưng tổng thể cho ảnh đầu vào.

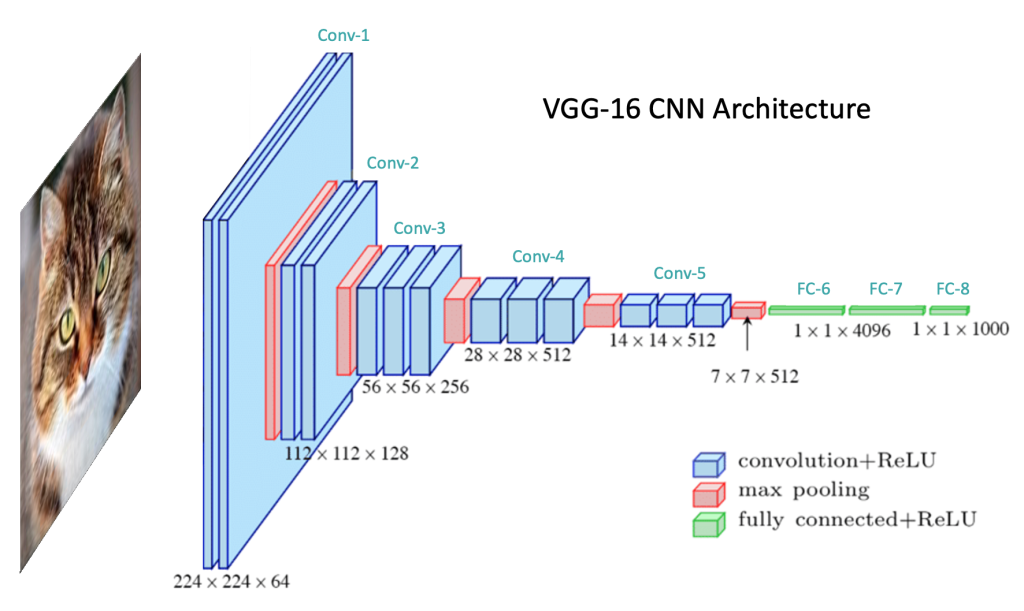
Mô hình CNN đã chứng tỏ được tính hiệu quả trong nhiều ứng dụng như phân loại ảnh, nhận dạng khuôn mặt và xử lý ngôn ngữ tự nhiên.

1. **Cấu trúc CNN**

Mô hình CNN (Convolutional Neural Network) là một kiến trúc mạng nơ-ron tích chập được sử dụng rộng rãi trong lĩnh vực thị giác máy tính. Mô hình này bao gồm các tầng tích chập và tầng tổng hợp (pooling) để trích xuất đặc trưng của hình ảnh, sau đó sử dụng các tầng đầy đủ kết nối (fully connected layer) để phân loại ảnh.

Một mô hình CNN cơ bản bao gồm các thành phần chính sau:

* *Tầng đầu vào (Input layer)*: Tầng này nhận dữ liệu hình ảnh đầu vào và chuyển tiếp cho tầng tiếp theo.
* *Tầng tích chập (Convolutional layer)*: Tầng này sử dụng các bộ lọc (filter) để trích xuất các đặc trưng của hình ảnh. Bộ lọc sẽ di chuyển qua toàn bộ hình ảnh và tính tích chập (convolution) giữa bộ lọc và các vùng trùng khớp của hình ảnh, tạo ra một bản đồ đặc trưng (feature map) mới.
* *Tầng kích hoạt (Activation layer)*: Tầng này áp dụng một hàm kích hoạt phi tuyến (nonlinear activation function) lên các giá trị trong bản đồ đặc trưng, giúp mô hình học được các đặc trưng phi tuyến của hình ảnh.
* *Tầng tổng hợp (Pooling layer)*: Tầng này giảm kích thước của bản đồ đặc trưng bằng cách chọn ra các giá trị lớn nhất (max pooling) hoặc trung bình (average pooling) từ các vùng không chồng lên nhau của bản đồ đặc trưng. Điều này giúp giảm số lượng tham số và tính toán trong mô hình, đồng thời giúp mô hình trở nên bền vững với các biến động nhỏ trong vị trí của các đặc trưng trong hình ảnh.
* *Tầng đầy đủ kết nối (Fully connected layer)*: Tầng này nhận đầu vào là một vector được tạo ra bằng cách làm phẳng bản đồ đặc trưng từ tầng trước đó và áp dụng một số tầng kết nối đầy đủ để phân loại ảnh. Cuối cùng, một tầng softmax được sử dụng để tính xác suất của các lớp phân loại, đưa ra dự đoán cuối cùng.
* *Tầng đầu ra (Output layer)*: Tầng này cho ra đầu ra cuối cùng của mô hình, tức là kết quả phân loại của hình ảnh. Tùy thuộc vào loại bài toán, tầng này có thể sử dụng các hàm kích hoạt khác nhau như hàm softmax cho bài toán phân loại đa lớp, hàm sigmoid cho bài toán phân loại nhị phân, hoặc không sử dụng hàm kích hoạt cho bài toán regression.



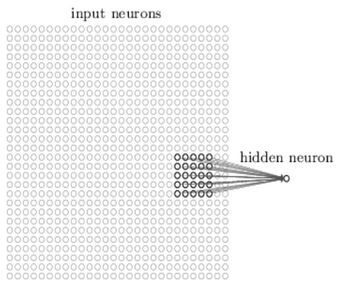
**Mạng CNN sử dụng 3 ý tưởng cơ bản:**

* Các trường tiếp nhận cục bộ (local receptive field)
* Trọng số chia sẻ (shared weights)
* Tổng hợp (pooling).
  1. *Trường tiếp nhận cục bộ (local receptive field)*

Trường tiếp nhận cục bộ (local receptive field) trong CNN được hiểu là kích thước của vùng trong hình ảnh mà một tầng tích chập của mô hình sẽ quét qua để trích xuất các đặc trưng. Vùng này được xác định bởi kích thước của bộ lọc (filter) và bước nhảy (stride) giữa các vùng liền kề.

Đầu vào của mạng CNN là một ảnh. Ví dụ như ảnh có kích thước 28×28 thì tương ứng đầu vào là một ma trận có 28×28 và giá trị mỗi điểm ảnh là một ô trong ma trận. Trong mô hình mạng ANN truyền thống thì chúng ta sẽ kết nối các neuron đầu vào vào tầng ảnh.

Tuy nhiên trong CNN chúng ta không làm như vậy mà chúng ta chỉ kết nối trong một vùng nhỏ của các neuron đầu vào như một filter có kích thước 5×5 tương ứng (28- 5 + 1) 24 điểm ảnh đầu vào. Mỗi một kết nối sẽ học một trọng số và mỗi neuron ẩn sẽ học một bias. Mỗi một vùng 5×5 đấy gọi là một trường tiếp nhận cục bộ.

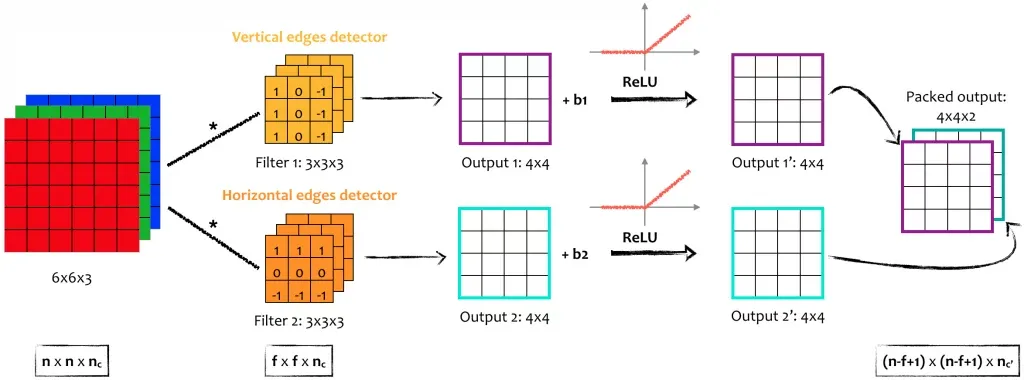


Một trong những ưu điểm của trường tiếp nhận cục bộ là nó cho phép mô hình tập trung vào các đặc trưng cục bộ trong hình ảnh, giúp giảm thiểu ảnh hưởng của các đối tượng không quan trọng trong hình ảnh và tăng tính bền vững của mô hình đối với các biến động trong đối tượng.

Tuy nhiên, một nhược điểm của trường tiếp nhận cục bộ là nó có thể dẫn đến mất mát thông tin toàn cục trong hình ảnh, do mô hình chỉ tập trung vào các đặc trưng cục bộ. Điều này có thể ảnh hưởng đến khả năng của mô hình trong việc phân loại các hình ảnh có cấu trúc phức tạp, chẳng hạn như hình ảnh có nhiều đối tượng chồng lên nhau hoặc hình ảnh có bối cảnh phức tạp.

* 1. *Trọng số chia sẻ (shared weight and bias)*

Trọng số chia sẻ (shared weight) là một tính năng của mô hình CNN, trong đó các trọng số của bộ lọc được sử dụng để trích xuất đặc trưng từ ảnh đầu vào được chia sẻ trong suốt quá trình tính toán của mạng. Điều này có nghĩa là bộ lọc được sử dụng để trích xuất đặc trưng từ một vùng của ảnh cũng sẽ được sử dụng để trích xuất đặc trưng từ các vùng khác trong ảnh.



Điều này giúp giảm số lượng tham số trong mô hình, đồng thời cải thiện khả năng tổng quát hóa của mô hình. Nếu các trọng số không được chia sẻ, mô hình sẽ có quá nhiều tham số, dẫn đến tình trạng quá khớp và kém hiệu quả trong việc tổng quát hóa trên các dữ liệu mới.

Trong khi đó, bias cũng được chia sẻ giữa các bộ lọc khác nhau, giúp đưa các giá trị tích chập về khoảng giá trị chính xác hơn.

* 1. *Tổng hợp (pooling)*

Tổng hợp (pooling) là một kỹ thuật quan trọng trong các mô hình CNN để giảm kích thước của bản đồ đặc trưng (feature map) và giảm số lượng tham số trong mô hình.

Có hai loại tổng hợp thường được sử dụng trong các mô hình CNN: Max pooling và average pooling.

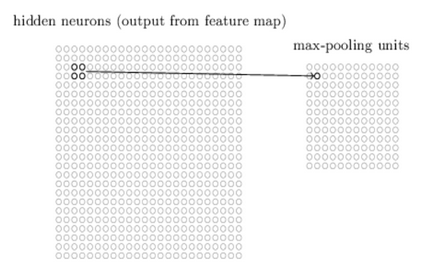
* + Max pooling: Kỹ thuật này chọn ra giá trị lớn nhất trong mỗi cửa sổ trượt (sliding window) trên bản đồ đặc trưng. Giá trị lớn nhất này sẽ được lấy để thay thế toàn bộ cửa sổ trượt. Max pooling giúp giảm kích thước của bản đồ đặc trưng và tạo ra bản đồ đặc trưng mới chỉ chứa các giá trị quan trọng nhất.
  + Average pooling: Kỹ thuật này tính giá trị trung bình của mỗi cửa sổ trượt trên bản đồ đặc trưng. Giá trị trung bình này sẽ được lấy để thay thế toàn bộ cửa sổ trượt. Average pooling cũng giúp giảm kích thước của bản đồ đặc trưng nhưng không giữ lại thông tin quan trọng nhất như max pooling.

Các lợi ích của kỹ thuật tổng hợp trong mô hình CNN bao gồm:

* Giảm kích thước của bản đồ đặc trưng, giúp giảm số lượng tham số và tính toán trong mô hình.
* Giúp mô hình trở nên bền vững với các biến động nhỏ trong vị trí của các đặc trưng trong hình ảnh.
* Giúp giảm overfitting trong mô hình, do giảm số lượng tham số và tính toán.



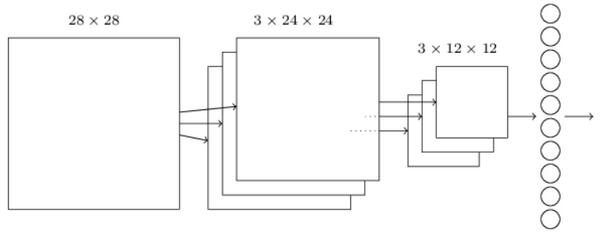
Hình 1. Lớp pooling thường được sử dụng ngay sau lớp convational để giảm bớt thông tin đầu ra



Hình 2. Thủ tục pooling phổ biến nhất là max-pooling

Như vậy qua lớp Max Pooling thì số lượng neuron giảm đi phân nửa. Trong một mạng CNN có nhiều Feature Map nên mỗi Feature Map chúng ta sẽ cho mỗi Max Pooling khác nhau. Chúng ta có thể thấy rằng Max Pooling là cách hỏi xem trong các đặc trưng này thì đặc trưng nào là đặc trưng nhất. Ngoài Max Pooling còn có L2 Pooling.

Cuối cùng ta đặt tất cả các lớp lại với nhau thành một CNN với đầu ra gồm các neuron với số lượng tùy bài toán.



Hai lớp cuối cùng của các kết nối trong mạng là một lớp đầy đủ kết nối (fully connected layer) . Lớp này nối mọi nơ-ron từ lớp max pooled tới mọi nơron của tầng ra.

1. **Tham số CNN**

Bên cạnh những tham số cơ bản như số lượng convolution layer, kích thước filter, kích thước pooling, còn có những tham số khác cũng ảnh hưởng đến hiệu suất của mô hình CNN, bao gồm:

* *Số lượng filter trong mỗi convolution layer*: Thông thường, số lượng filter sẽ tăng dần từ layer đầu tiên đến layer cuối cùng. Tuy nhiên, quá nhiều filter cũng có thể dẫn đến overfitting và tăng độ phức tạp tính toán.
* *Hàm kích hoạt trong mỗi layer*: Ngoài hàm ReLU thường được sử dụng, còn có các hàm kích hoạt khác như sigmoid, tanh, leaky ReLU, ELU... Tùy vào bài toán cụ thể, ta có thể chọn hàm kích hoạt phù hợp để cải thiện hiệu suất của mô hình.
* *Dropout*: Là một kỹ thuật regularization giúp tránh overfitting. Dropout giúp ngẫu nhiên tắt một số node trong mỗi lần huấn luyện để giảm độ phức tạp của mô hình.
* *Learning rate*: Là tham số quan trọng trong quá trình huấn luyện mô hình. Tốc độ học tập quá cao có thể dẫn đến overshooting, trong khi tốc độ học tập quá thấp có thể dẫn đến hội tụ chậm hoặc dễ bị kẹt ở local minimum.
* *Regularization*: Có nhiều kỹ thuật regularization khác nhau như L1 regularization, L2 regularization, early stopping... giúp tránh overfitting và cải thiện khả năng tổng quát hóa của mô hình.
* *Kích thước batch*: Là số lượng mẫu được sử dụng để cập nhật trọng số trong mỗi lần huấn luyện. Kích thước batch quá nhỏ có thể làm cho quá trình huấn luyện chậm, trong khi kích thước batch quá lớn có thể dẫn đến khó khăn trong việc tối ưu hóa mô hình.

Các tham số này phải được điều chỉnh cẩn thận để đạt được hiệu suất tốt nhất cho mô hình CNN.

Cách chọn tham số cho mô hình CNN có thể được thực hiện bằng cách sử dụng các kỹ thuật tuning và regularization, bao gồm:

* Tinh chỉnh learning rate: Đây là một tham số quan trọng của thuật toán tối ưu hóa sử dụng trong CNN. Nếu learning rate quá lớn, có thể dẫn đến sự dao động của mô hình và khó khắc phục. Nếu learning rate quá nhỏ, thì mô hình sẽ hội tụ chậm hơn. Vì vậy, cần phải tìm kiếm giá trị learning rate tối ưu thông qua các kỹ thuật tuning, chẳng hạn như sử dụng lược đồ tìm kiếm mạng.
* Sử dụng các kỹ thuật regularization: Các kỹ thuật regularization như dropout, L1 regularization và L2 regularization có thể được sử dụng để giảm overfitting của mô hình. Các kỹ thuật này giúp mô hình tránh được việc nhớ các điểm dữ liệu trong tập huấn luyện và dẫn đến hiệu suất tốt hơn trên tập kiểm tra.
* Tinh chỉnh kích thước batch: Kích thước batch là số lượng mẫu được đưa vào mô hình để huấn luyện trong mỗi lần cập nhật trọng số. Nếu kích thước batch quá lớn, sẽ dẫn đến tốc độ huấn luyện chậm và yêu cầu bộ nhớ lớn hơn. Nếu kích thước batch quá nhỏ, thì gradient ước lượng của mô hình sẽ không chính xác, làm giảm hiệu suất của mô hình. Vì vậy, cần tìm kiếm kích thước batch tối ưu để tối ưu hiệu suất và tốc độ huấn luyện của mô hình.
* Tinh chỉnh số lượng tầng và số lượng bộ lọc: Số lượng tầng và số lượng bộ lọc là các tham số quan trọng trong mô hình CNN. Số lượng tầng và số lượng bộ lọc khác nhau có thể tạo ra các mô hình khác nhau với hiệu suất và tốc độ huấn luyện khác nhau. Vì vậy, cần phải tìm kiếm số lượng tầng và số lượng bộ lọc tối ưu để tối ưu hiệu suất của mô hình.

1. **Ưu, nhược điểm của thuật toán CNN**

*3.1. Một số ưu điểm của thuật toán CNN*

* *Tính tổng quát cao*: Thuật toán CNN có khả năng học và trích xuất các đặc trưng cục bộ của hình ảnh một cách tự động, giúp mô hình có khả năng nhận diện các đối tượng trên hình ảnh ở nhiều vị trí khác nhau. Do đó, mô hình CNN có khả năng tổng quát hóa tốt hơn so với các thuậttoán học máy truyền thống.
* *Hiệu suất cao*: Thuật toán CNN có thể xử lý một lượng lớn dữ liệu hình ảnh và trích xuất các đặc trưng từ chúng trong thời gian ngắn. Điều này giúp mô hình CNN có hiệu suất cao và đáp ứng được yêu cầu của các ứng dụng thời gian thực như xử lý video và hình ảnh.
* *Dễ dàng mở rộng*: Mô hình CNN có thể dễ dàng mở rộng bằng cách thêm hoặc loại bỏ các tầng khác nhau tùy vào yêu cầu của bài toán. Điều này giúp mô hình CNN linh hoạt và có thể được sử dụng trong nhiều ứng dụng khác nhau.
* *Tính tái sử dụng cao*: Các bộ lọc (filter) và tầng tích chập trong thuật toán CNN có thể được sử dụng lại trong nhiều bài toán khác nhau, giúp tiết kiệm thời gian và tài nguyên trong quá trình phát triển mô hình.
* *Tính phi tuyến cao*: Các tầng kích hoạt phi tuyến trong mô hình CNN giúp mô hình học được các đặc trưng phi tuyến của hình ảnh, giúp nó có khả năng phân loại chính xác các hình ảnh phức tạp.
  1. *Một số nhược điểm của thuật toán CNN*

Mặc dù CNN có nhiều ưu điểm, nhưng cũng có một số nhược điểm như sau:

* *Tính đối xứng của bộ lọc*: Bộ lọc trong CNN là đối xứng, có nghĩa là chúng cũng có thể tìm kiếm các đặc trưng ở những vị trí đối xứng trên hình ảnh, điều này làm giảm khả năng phân biệt giữa các đối tượng đối xứng trên hình ảnh.
* *Overfitting*: CNN có thể dễ dàng bị overfitting với dữ liệu đào tạo, đặc biệt là khi số lượng mẫu huấn luyện quá nhỏ hoặc các bộ lọc quá phức tạp.
* *Cần nhiều tài nguyên tính toán*: Với mô hình CNN có nhiều tầng và bộ lọc phức tạp, cần nhiều tài nguyên tính toán để huấn luyện và dự đoán.
* *Thiếu khả năng tổng quát hóa*: Mô hình CNN có thể dễ dàng học được các đặc trưng cục bộ trên các hình ảnh, nhưng nó có thể thiếu khả năng tổng quát hóa cho các hình ảnh mới nằm ngoài phạm vi của dữ liệu huấn luyện.