**Thuyết Trình Máy Học**

Lời nói Đầu tiên xin gửi lời chào đến cô và các bạn, Cảm ơn mọi người đã dành thời gian để tham gia buổi thuyết trình hôm nay. Hôm nay, chủ đề thuyết trình của nhóm mình sẽ là nghiên cứu về Mạng Nơ-ron Tích Chập, từ viết tắt của nó là (CNNs) và cũng là CNN. Không làm phiền mọi người thêm nữa, chúng ta bắt đầu ngay.

Slide 2: Thành Viên Nhóm

Xin được giới thiệu nhóm mình gồm có ba thành viên

Người thứ nhất đó là mình, tiếp đến là bạn Huỳnh Nguyễn Thanh Trúc cuối cùng là bạn Lê Phước Hữu

Và sau đó chúng ta sẽ tìm hiểu xem chủ đề hôm nay gồm có những nội dung chính gì

Slide 3: Nội dung chính

Nội dung trình của bài gồm có 6 phần :

1 - Đầu tiên sẽ là lí do để chọn đề tài này là gì?

2 - Tiếp đến là đi vào phần tổng quan của Mạng Nơ-ron Tích Chập,Mạng nơ ron tích chập gồm những gì

3 - Sau đó là các ứng dụng của mạng nơ ron tích chập

4 - Tiếp đến là các bước để huấn luyện mô hình như thế nào

5 - Cuối cùng sẽ có 1 phần demo code liên quan đến nghiên cứu của mạng nơ ron tích chập

6 – Slide cuối cùng sẽ là danh mục các tài liệu mà mình tham khảo để nghiên cứu về đề tài này

Chúng ta sẽ vào phần đầu tiên lí do mà nhóm mình chọn đề tài này là gì.

Slide 4: Lí do chọn đề tài

Có 8 lí do để nhóm mình chọn đề tài nghiên cứu này

Thứ nhất : Hiệu suất xử lý hình ảnh của CNN cao

Thứ hai: CNN phổ biển và phát triển mạnh mẽ

Thứ ba: Đối tượng nghiên cứu phổ quát

Thứ tư: CNN được ứng dụng rộng rãi

Thứ năm: Gỉai quyết được các vấn đề thực tế

Thứ sáu: Chấp nhận được các dữ liệu đa dạng

Thứ bảy: Có thể phát triển được các kỹ năng chuyên sâu

Và cuối cùng là theo sở thích của nhóm mình

Mình sẽ đi vào phần quan trọng của bài đó là Tổng quan về mạng nơ ron tích chập CNN

Slide 5: Tổng quan về mạng nơ ron tích chập

Mạng nơ ron tích chập hay còn được gọi là CNN hoặc CNNs, đây là từ viết tắt của Convolutional Neural Network là một trong những mô hình Deep Learning cực kỳ tiên tiến, bởi vì chúng cho phép mình xây dựng những hệ thống có độ chính xác cao và thông minh,Nhờ vào khả năng đó CNN đã có rất nhiều ứng dụng và đặc biệt nhất là bài toán nhận dạng vật thể trong hình ảnh.

Mạng nơ ron tích chập là một loại mạng được thiết kế đặt biệt để xử lý các dữ liệu không gian như hình ảnh. CNN vô cùng quan trọng để tạo nên những hệ thống nhận diện thông minh với độ chính xác cao

(Gọi là CNNs hay CNN thì còn tuỳ thuộc vào ngữ cảnh ,ngữ nghĩa khác nhau

nếu

gọi là CNNs thì khi muốn nói về các mô hình nơ-ron có cấu trúc tích chập và được thiết kế đặc biệt để xử lý dữ liệu không gian như hình ảnh. Còn từ "CNN" thì đôi khi được sử dụng một cách tổng quát hơn để chỉ mạng nơ-ron sử dụng tích chập không riêng biệt cho việc xử lý hình ảnh.)

Mình sẽ đi đến phần tiếp theo của CNN đó chính là các lớp cơ bản của mạng CNN

Slide 6: Các lớp cơ bản của mạng CNN

CNN gồm có 4 lớp cơ bản:

* Thứ nhất là lớp tích chập (Convolutional Layer) *Lớp này giúp học các đặc trưng từ dữ liệu không gian, như hình ảnh. Nó sử dụng các bộ lọc (kernels) để thực hiện phép tích chập trên dữ liệu đầu vào.*
* Lớp thứ hai là lớp gộp (Pooling Layer) *Lớp này giúp giảm kích thước của biểu diễn không gian của dữ liệu và giảm chiều sâu của đầu ra. Phổ biến nhất là lớp gộp bản địa (max pooling) và lớp gộp trung bình (average pooling).*
* Lớp thứ ba là lớp kích hoạt (Activation Layer) *Lớp này thường là một hàm kích hoạt hàm đó thường là hàm ReLU được áp dụng để tăng tính phi tuyến tính của mô hình.*
* Lớp thứ tư là lớp kết nối đầy đủ (Fully Connected Layer) *Lớp này kết nối mỗi nơ-ron với mọi nơ-ron trong lớp trước đó, thường xuất hiện ở cuối của mô hình để tạo ra đầu ra phân loại.*

Chúng ta sẽ đi sâu vào các lớp của CNN đầu tiên là lớp tích chập

Slide 7: Lớp tích chập

*Lớp này giúp học các đặc trưng từ dữ liệu không gian, như hình ảnh. Nó sử dụng các bộ lọc (kernels) để thực hiện phép tích chập trên dữ liệu đầu vào.*

* Convolutional chính là một loại cửa sổ dạng trượt nằm trên một ma trận. Các convolutional layer sẽ chứa các parameter có khả năng tự học, qua đó thì sẽ điều chỉnh và tìm ra cách lấy những thông tin chính xác nhất trong khi không cần chọn feature.

Lúc này thì tích chập đóng vai trò là nhân các phần tử thuộc ma trận. Sliding Window, hay được gọi là kernel, filter hoặc feature detect, là loại ma trận có kích thước nhỏ.

Slide 8

* Đây chính là lớp đóng vai trò mấu chốt của CNN, khi layer này đảm nhiệm việc thực hiện mọi tính toán. Stride, padding, filter map, feature map là những yếu tố quan trọng nhất của convolutional layer. Convolutional Layer chịu trách nhiệm trích xuất đặc trưng từ ảnh.

Có khả năng nhận diện các đặc trưng không gian như cạnh, góc, và các đặc điểm phức tạp hơn.

* Bộ lọc (kernel) sẽ di chuyển qua ảnh để tìm các đặc trưng cụ thể.
* Mỗi bộ lọc tạo ra một ma trận đặc trưng (feature map).
* Stride: là sự dịch chuyển filter map theo pixel dựa trên giá trị từ trái sang phải.
* Padding: Là các giá trị 0 được thêm cùng lớp input.
* Feature map: Sau mỗi lần quét, một quá trình tính toán sẽ được thực hiện. Feature map sẽ thể hiện kết quả sau mỗi lần filter map quét qua input.
* Ứng dụng hàm kích hoạt để bắt cặp các đặc trưng.

Tiếp đến sẽ là lớp gộp

Slide 9: Lớp gộp

*Lớp này giúp giảm kích thước của biểu diễn không gian của dữ liệu và giảm chiều sâu của đầu ra. Phổ biến nhất là lớp gộp bản địa (max pooling) và lớp gộp trung bình (average pooling).* *Lớp pooling sẽ giảm bớt số lượng tham số khi hình ảnh quá lớn. Không gian pooling còn được gọi là lấy mẫu con hoặc lấy mẫu xuống làm giảm kích thước của mỗi map nhưng vẫn giữ lại thông tin quan trọng.*

***Average Pooling:****Quá trình average pooling thực hiện bằng cách chia input feature map thành các ô không chồng lên nhau và tính giá trị trung bình của các giá trị trong mỗi ô. Kết quả là một feature map mới với kích thước giảm đi, nhưng vẫn giữ các đặc trưng quan trọng của hình ảnh*

Lớp gộp giúp giảm kích thước của đặc trưng. Loại bỏ thông tin không quan trọng, giữ lại những đặc trưng quan trọng. Trích xuất thông tin quan trọng từ các vùng hình ảnh.

* Max Pooling chọn giá trị lớn nhất từ mỗi vùng, Average Pooling lấy giá trị trung bình. đòi hỏi phải quét qua một hình ảnh bằng cách sử dụng bộ lọc và tại mỗi trường hợp, trả về giá trị pixel tối đa bị bắt trong bộ lọc dưới dạng pixel của riêng nó trong một hình ảnh mới.
* Giảm lượng tính toán và tăng tính chia sẻ đặc trưng.

Slide 10: Lớp kích hoạt

*Lớp này thường là một hàm kích hoạt hàm đó thường là hàm ReLU được áp dụng để tăng tính phi tuyến tính của mô hình.*

- Lớp kích hoạt (Activation Layer) trong mạng nơ-ron là một phần quan trọng giúp tăng tính phi tuyến tính của mô hình. Mục tiêu chính của lớp này là thêm không gian phi tuyến tính vào đầu ra của tầng trước đó, giúp mô hình có khả năng học được các đặc trưng phức tạp và biểu diễn không gian không gian. thường là một hàm kích hoạt (ReLU)

- ReLU là viết tắt của Rectified Linear Unit là một trong những hàm kích hoạt phổ biến được sử dụng trong lớp kích hoạt. Hàm ReLU đơn giản là xác định giá trị đầu ra là max(0, x), nghĩa là nếu giá trị đầu vào là dương, giữ nguyên giá trị đó, còn nếu giá trị đầu vào là âm, đầu ra sẽ là 0.

f(x)=max(0,x)

X : giá trị đầu vào

F(x): Giá trị đầu ra của hàm ReLu

Và lớp cuối cùng đó chính là lớp kết nối đầy đủ

Slide 11: Lớp kết nối đầy đủ

*Tầng kết nối đầy đủ là một loại tầng nhận các dữ liệu đầu vào đã dược làm phẳng để kết nối mỗi nơ-ron đầu ra của tầng trước đó với mỗi nơ-ron đầu vào của tầng đó. Trong tầng này, mỗi đầu vào được kết nối với tất cả các đầu ra thông qua các trọng số (weights).*

Lớp này thường sẽ xuất hiện ở các lớp cuối cùng của mô hình.

Kết nối mỗi neuron của layer trước với mỗi neuron của layer hiện tại.

Thực hiện phân loại dựa trên đặc trưng đã được học.

Kích thước của layer này phụ thuộc vào số lượng neurons và lớp cần phân loại.

Liên kết các nơ-ron với nhau để tạo ra dự đoán cuối cùng.

Các nơ-ron đầy đủ kết nối để phân loại và dự đoán đầu ra.

Slide 12: Ứng dụng

CNN có rất nhiều ứng dụng trong nhiều lĩnh vực khác nhau và dưới đây là một số ứng dụng phổ biến như phân loại ảnh, phát hiện đối tượng, nhận diện khuôn mặt

* Phân Loại Ảnh: Hệ thống nhận diện và phân loại các đối tượng hoặc phân loại hình ảnh thành các danh mục cụ thể.

Ví Dụ: Dịch vụ phân loại nội dung, nhận diện sản phẩm trong thương mại điện tử, phân loại loại cây trong ảnh,…

* Phát Hiện Đối Tượng: Tìm kiếm và xác định vị trí của các đối tượng trong hình ảnh.

Ví Dụ: Hệ thống an ninh nhận diện vật thể nghi ngờ, hệ thống giám sát giao thông phát hiện xe cản trở, công nghệ giám sát sản xuất để kiểm tra sản phẩm,…

* Nhận Diện Khuôn Mặt: Xác định và nhận diện các khuôn mặt trong hình ảnh hoặc video.

Ví Dụ: Hệ thống mở khóa điện thoại di động bằng nhận diện khuôn mặt,...

Slide 13: Ưu điểm

Sau khi mình tìm hiểu về CNN thì kết luận được CNN có 4 ưu điểm

* Khả Năng Học Đặc Trưng Tự Động: CNN có khả năng tự học các đặc trưng từ dữ liệu, giảm công sức cần thiết cho việc thiết kế và chọn lọc đặc trưng thủ công.
* Hiệu Suất Cao Trong Xử Lý Hình Ảnh: CNN được thiết kế đặc biệt cho việc xử lý dữ liệu không gian như hình ảnh và video, với khả năng nhận diện đối tượng và cấu trúc không gian.
* Tích Hợp Tốt với Dữ Liệu Lớn: Hiệu suất của CNN thường tăng lên đáng kể khi được huấn luyện trên dữ liệu lớn, giúp cải thiện khả năng tổng quát hóa của mô hình.
* Chia Sẻ Trọng Số (Weight Sharing): CNN sử dụng cơ chế chia sẻ trọng số giữa các vùng không gian, giúp giảm số lượng tham số cần học và giảm nguy cơ quá mức.

Slide 14: Nhược điểm

* Yêu Cầu Nguồn Lực Lớn: Huấn luyện CNN đòi hỏi nguồn lực tính toán và bộ nhớ lớn, đặc biệt là với các mô hình sâu và dữ liệu lớn.
* Cần Dữ Liệu Lớn: Để đạt hiệu suất tốt, CNN thường yêu cầu một lượng lớn dữ liệu huấn luyện, điều này có thể là một rào cản đối với các ứng dụng có dữ liệu hạn chế.
* Không Hiệu Quả Trong Việc Xử Lý Dữ Liệu Tuần Tự: CNN không hiệu quả khi xử lý dữ liệu có tính tuần tự, chẳng hạn như trong ngôn ngữ tự nhiên. Các mô hình RNNs thường được ưa chuộng hơn trong trường hợp này.
* Dễ Bị Quá Mức (Overfitting): Trong trường hợp dữ liệu huấn luyện không đủ hoặc mô hình quá phức tạp, có nguy cơ mô hình bị quá mức (overfitting) và không tổng quát hóa tốt trên dữ liệu mới.

Slide 15: Qúa trình huấn luyện mô hình để demo

Bước 1: Chuẩn bị dữ liệu: Thu thập dữ liệu và tiền xử lý dữ liệu chuẩn bị dữ liệu gồm có các tập để huấn luyện và các tập dữ liệu để test kiểm tra

Bước 2: Chuẩn hóa dữ liệu: Chuyển đổi giá trị pixel về khoảng [0, 1] hoặc [-1, 1].

Định dạng lại kích thước ảnh nếu cần thiết để phù hợp với kiến trúc mô hình.

Bước 3: Huấn luyện mô hình: Chia dữ liệu Huấn Luyện và test

Sử dụng dữ liệu huấn luyện để điều chỉnh trọng số của mô hình theo hướng làm giảm giá trị hàm mất mát. Sau đó Theo Dõi và Đánh Giá

Bước 4: Đánh Giá và Triển Khai:

* Đánh Giá Mô Hình: Sử dụng tập kiểm thử để đánh giá hiệu suất cuối cùng của mô hình. Đánh giá mô hình trên dữ liệu mới nếu có.
* Triển Khai: Chuyển mô hình từ môi trường phát triển sang môi trường triển khai.

Xử lý các vấn đề hiệu suất và tối ưu hóa nếu cần thiết để tích hợp mô hình vào ứng dụng thực tế.

Và sau đây là phần demo code của mình phần demo này mình đã sử dụng mô hình YOLOv8 để

NHẬN DIỆN CON NGƯỜI VÀ QUẢ BÓNG

Tập dữ liệu gồm 59 ảnh và 59 label đã được gắn nhãn bằng web Roboflow

Mô hình đã được train 5 lần và đã test thử trên các ảnh và mình sẽ chọn ngẫu nhiên ảnh số 5 để nhận diện thử cho mọi người xem

• Tập dữ liệu (Dataset)

o Folder <images>

59 ảnh

o Folder <labels>

59 label

• Training

Đã train sẵn 5 lần

• Test

Nhận diện ảnh số 5

Cuối cùng là phần tài liệu mà mình đã tham khảo để nghiên cứu về mạng nơ ron tích chập

Slide cảm ơn:

Và bài thuyết trình của mình đến đây là kết thúc xin Trân trọng cảm ơn mọi người đã lắng nghe!