ĐẠI HỌC BÁCH KHOA HÀ NỘI

**TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

----- 🙡 🕮 🙣 -----



**PROJECT I**

***Đề tài:* Tìm hiểu mạng neuron tích chập. Ứng dụng mạng nơ ron tích chập trong bài toán phân loại ảnh.**

Giảng viên: **Ths. Lê Đức Trung**

Sinh viên thực hiện: Vũ Văn Đức

Mã sinh viên: 20194025

**Hà Nội, năm 2021**

Mục lục

|  |  |
| --- | --- |
| Tên mục | Trang |
| Mở đầu | 3 |
| Chương 1 Tổng qua về mạng nơ ron nhân tạo và giới thiệu mạng nơ ron tích chập | 4 |
| I.1 Tổng quan mạng nơ ron nhân tạo | 4 |
| I.1.1 Lịch sử mạng nơ ron | 4 |
| I.1.2 Mạng nơ ron sinh học | 4 |
| I.1.3 Mạng nơ ron nhân tạo | 5 |
| I.1.4 Các hàm kích hoạt | 5 |
| I.2 Mạng nơ ron nhân tạo | 8 |
| I.2.1 Giới thiệu mạng nơ ron nhân tạo | 8 |
| I.2.2 Một số mạng nơ ron nhân tạo | 8 |
| I.2.3 Thuật toán lan truyền ngược | 9 |
| I.3 Mạng nơ ron tích chập | 10 |
| I.3.1 Giới thiệu mạng nơ ron tích chập | 11 |
| I.3.2 Mô hình mạng nơ ron tích chập | 11 |
| I.3.3 Xây dựng mạng nơ ron tích chập | 11 |
| I.4 Kết luận | 14 |
| Chương 2 Bài toán phân loại hình ảnh bằng mạng nơ ron tích chập | 16 |
| II.1 Giới thiệu bài toán | 16 |
| II.2 Thực nghiệm | 17 |
| II.2.1 Mô hình và tập dữ liệu | 17 |
| II.2.2 Kết quả đạt được | 19 |
| II.2.3 Hướng phát triển trong tương lai | 20 |
| Kết luận | 20 |
| Tài liệu tham khảo | 22 |

**MỞ ĐẦU**

**Lý do chọn đề tài**

Hiện nay, máy tính có sức mạnh tính toán lớn nhưng giá thành lại ở mức phổ thông, dẫn tới người làm nghiên cứu rất dễ dàng để có thể tự kiểm nghiệm được các lý thuyết về trí tuệ nhân tạo từ nhiều năm trước. Cùng với mã nguồn mở, hiện làn sóng trí tuệ nhân tạo đang bùng nổ mạnh mẽ trong thời gian gần đây, và đem lại rất nhiều ứng dụng trong đời sống.

Chính vì lý do trên, việc tìm hiểu và nghiên cứu thị giác máy tính rất có ý nghĩa và thiết thực. Hiện nay trên thế giới, nhiều nước đã và đang ứng dụng thị giác máy tính vào đời sống hàng ngày, ví dụ như SkyNet của Trung Quốc, các hệ thống kiểm duyệt nội dung tự động, …

**Cơ sở khoa học**

Ngày nay, máy tính là một phần không thể thiếu trong nghiên cứu khoa học cũng như trong đời sống hàng ngày. Tuy nhiên, do hệ thống máy tính dựa trên lý thuyết cổ điển (tập hợp, logic nhị phân), nên dù có khả năng tính toán lớn và độ chính xác cao, thì máy tính cũng chỉ có thể làm việc theo một chương trình gồm các thuật toán được viết sẵn do lập trình viên chứ chưa thể tự lập luận hay sáng tạo.

Chính vì lý do đó, học máy (machine learning) – một phương pháp phân tích dữ liệu mới ra đời, có thể tự động hóa phân tích dữ liệu và trích xuất các đặc trưng nằm sâu trong dữ liệu mà không thể lập trình bằng cách thông thường.

Học sâu (deep learning) là một nhánh đặc biệt của machine learning. Học sâu sử dụng mạng nơ ron nhân tạo (Artificial Neural Networks) để giải quyết các bài toán mà tính toán cứng chưa thể xử lý được như dữ liệu trừu tượng, dữ liệu mờ, dữ liệu không rõ ràng,... Với tính chất như vậy, học sâu thường được ứng dụng trong nhận diện hình ảnh, nhận diện giọng nói, xử lý ngôn ngữ tự nhiên hoặc dự đoán, dự báo.

Mạng nơ ron Tích chập (Convolutional Neural Network) là một trong những mô hình học sâu hiện đại nhất hiện nay. Mạng nơ ron Tích chập hiện nay thường được sử dụng nhiều trong các hệ thống thông minh do ưu điểm của mạng là có độ chính xác cao, tuy nhiên tốc độ tính toán lại rất nhanh.

Vì vậy trong đồ án môn học này, em tìm hiểu về cấu tạo, cách hoạt động của một mạng neuron tích chập và ứng dụng trong bài toán phân loại hình ảnh với mô hình VGG16 và tập dữ liệu ảnh hoa quả với 30 loại khác nhau. Sau khi huấn luyện mô hình dự đoán khá hiệu quả với độ chính xác trên tập đánh giá đạt 99,46% và trên tập test đạt 99,46%.

**CHƯƠNG I**

**TỔNG QUAN VỀ MẠNG NƠ RON VÀ GIỚI THIỆU VỀ MẠNG NƠ RON TÍCH CHẬP**

**I.1. Tổng quan mạng nơ ron nhân tạo**

**I.1.1. Lịch sử nơ ron nhân tạo**

Mạng nơ ron nhân tạo hay thường gọi là mạng neuron là một mô hình toán học hay mô hình tính toán được xây dựng dựa trên ý tưởng của mạng neuron sinh học.

Vào năm 1943, nhà thần kinh học Warren McCulloch đã cùng nhà toán học Walter Pitts đã viết một cuốn sách về cách mạng thần kinh hoạt động. Và họ đã thực hiện mô phỏng một mạng thần kinh đơn giản trên một mạch điện.

Vào năm 1949, Donald Hebb đã viết cuốn sách Organization of Behavior. Điểm nhấn chính là mạng thần kinh nào được sử dụng nhiều sẽ được tăng cường.

Vào năm 1959, David Hubel và Torsten Wiesel đã xuất bản cuốn sách Receptive fields of single neurons in the cat’s striate cortex, miêu tả về phản ứng của các tế bào thần kinh thị giác trên loài mèo, cũng như cách loài mèo ghi nhớ và nhận diện hình dạng trên kiến trúc vỏ não của nó.

Vào năm 1989, Yann LeCun đã áp dụng thuật toán học cho mạng nơ ron theo kiểu lan truyền ngược vào kiến trúc mạng nơ ron tích chập của Fukushima. Sau đó vài năm, LeCun đã công bố LeNet-5.

**I.1.2. Mạng nơ ron sinh học**

Một nơ ron gồm có: thân nơ ron, tua gai thần kinh, sợi trục thần kinh, trong đó:

• Thân nơ ron: là nơi xử lý các tín hiệu được đưa vào;

• Tua gai thần kinh: là nơi nhận các xung điện vào trong nơ ron;

• Sợi trục thần kinh: là nơi đưa tín hiệu ra ngoài sau khi được xử lý bởi nơ ron;

• Khớp thần kinh: vị trí nằm giữa tua gai thần kinh và sợi trục thần kinh, đây là điểm liên kết đầu ra của nơ ron này với đầu vào của nơ ron khác.

A close-up of a spider

Description automatically generated with medium confidence

Hình ảnh một nơ ron sinh học

**I.1.3 Mạng nơ ron nhân tạo.**

Dựa vào cấu tạo của một nơ ron sinh học, các nhà khoa học nghiên cứu và lập trình đã đưa ra kiến trúc của một nơ ron nhân tạo:

Diagram

Description automatically generated

**Hình ảnh cấu tạo một mạng nơ ron nhân tạo**

Trong đó :

• Danh sách các đầu vào: Là các thuộc tính đầu vào của một nơ ron. Số lượng thuộc tính đầu vào thường nhiều hơn một, do dữ liệu thô đầu vào thường là một vector nhiều chiều, hoặc nhiều nơ ron tầng trước kết nối tới một nơ ron tầng sau.

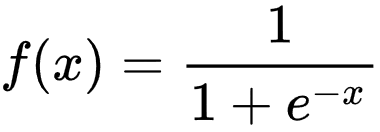
• Trọng số liên kết: Các liên kết được thể hiện độ mạnh yếu qua một giá trị được gọi là trọng số liên kết. Kết hơp với các đầu truyền, tín hiệu đến các nơ ron nhân tạo khác sẽ được tính bằng wi\*xi.

• Hàm tổng: Tổng các tích của các đầu vào với trọng số liên kết mô phỏng các khớp kết nối. Sau đó đi qua hàm tính tổng để tính ra giá trị trước khi đưa vào hàm kích hoạt.

• Hàm kích hoạt: Hàm kích hoạt của một nút định nghĩa đầu ra của nút đó được cho bởi một đầu vào hay tập đầu vào. Hiểu theo cách khác, với một đầu vào hay tập đầu vào, hàm kích hoạt sẽ cho ra đầu ra của một nút bất kỳ nào đó.

**I.1.4 Các hàm kích hoạt của một mạng nơ ron nhân tạo**

**I.1.4.1 Hàm sigmoid**

• Biểu diễn hàm: ****

• Đạo hàm: f’(x)=f(x)\*(1-f(x))

Hàm sigmoid được sử dụng vì ngưỡng của nó nằm trong khoảng (0, 1). Do đó, hàm này được sử dụng nhiều cho các mô hình dự đoán xác suất đầu ra, tức kết quả chỉ tồn tại trong khoảng từ 0 đến 1: khi đầu vào là số dương lớn, đầu ra của hàm sigmoid gần bằng 1. Khi nhỏ hơn 0, đầu ra gần bằng 0.

**Chart, histogram

Description automatically generated**

**Hình ảnh hàm sigmoid**

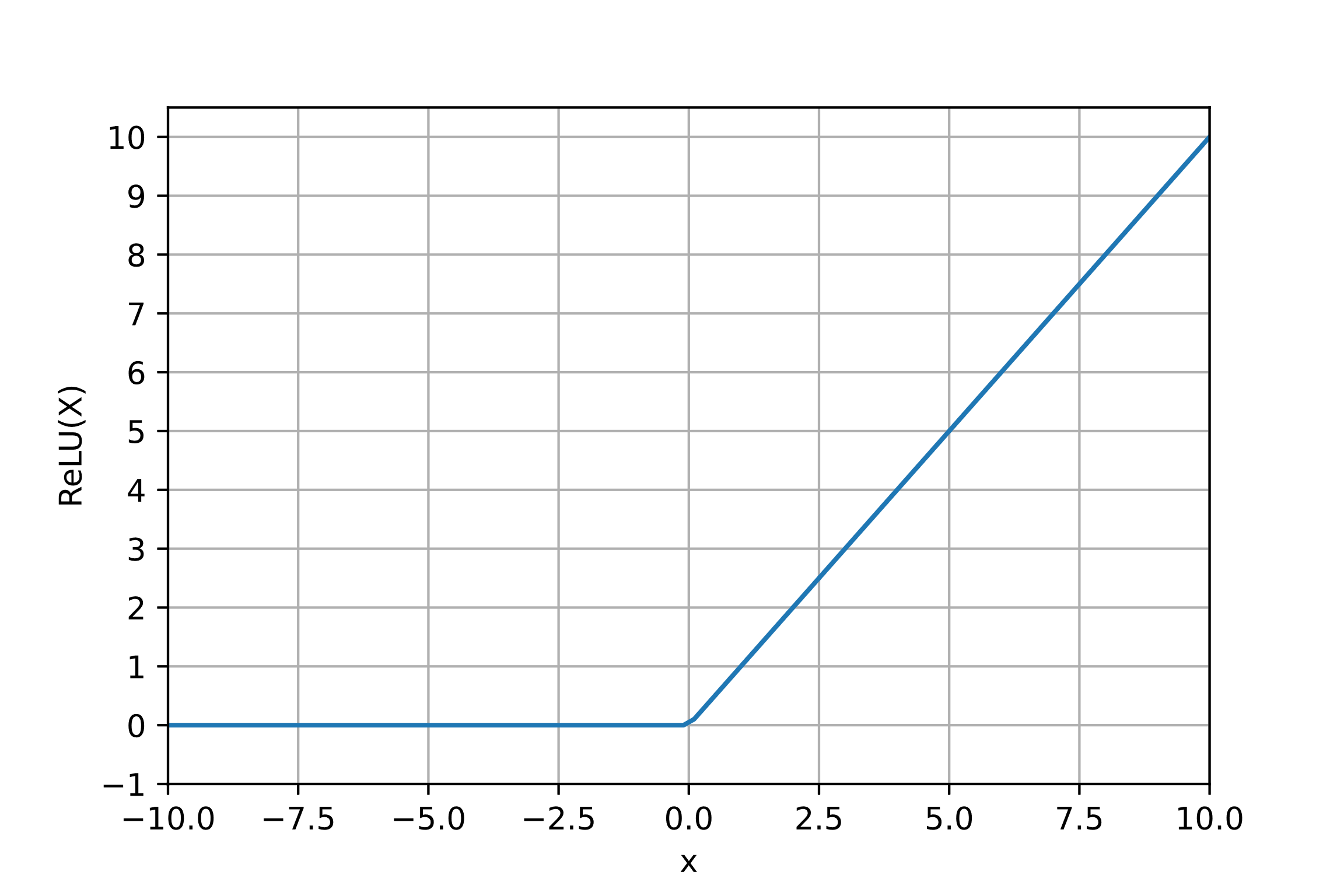
**I.1.4.2. Hàm Relu**

• Biểu diễn hàm: f(x)=max(0,x)

• Đạo hàm: f’(x)= 0 nếu x<0

1 nếu x>0

Hàm Relu áp dụng với những trường hợp cần đầu ra nằm trong khoảng (0, +∞). Hàm RELU có tốc độ tính toán rất nhanh, gán các giá trị âm trở thành 0 ngay lập tức, phù 10 hợp cho việc huấn luyện từ dữ liệu chuẩn. Tuy nhiên, điều này khiến hàm RELU không ánh xạ các giá trị âm một cách thích hợp.



**Hình ảnh hàm Relu**

**I.1.4.3 Hàm tanh**

• Biểu diễn hàm:

• Đạo hàm:

Hàm tanh được sử dụng vì đầu ra của hàm nằm trong khoảng , thích hợp với các mô hình đầu ra có ba giá trị: âm, trung tính (0) và dương.

**Diagram

Description automatically generated**

**Hình ảnh hàm tanh**

**I.1.4.4 Hàm softmax**

• Biểu diễn hàm:

• Đạo hàm:

Hàm softmax nhận đầu vào là một vec-tơ và cho đầu ra là 1 vec-tơ có cùng số chiều

Do với mọi x nên với mọi

Dễ thấy là hàm đồng biến mà mẫu số lại không đổi nên cũng là hàm đồng biến. Tức là với thì < . Lại có =1.

Từ những điều trên, ta liên tưởng tới xác suất:

Trong đó y là nhãn của tập dữ liệu x, là 1 nhãn của x.

A picture containing text, clock

Description automatically generated

**Hình ảnh minh họa giá trị của x sau khi đi qua hàm softmax**

**I.2. Mạng nơ ron nhân tạo**

**I.2.1. Giới thiệu mạng nơ ron nhân tạo**

Mạng nơ ron nhân tạo (Artificial Neural Network ANN) là một chuỗi các giải thuật lập trình, mô phỏng dựa trên cách hoạt động của mạng lưới thần kinh trong não bộ các sinh vật sống. Mạng nơ ron nhân tạo được sử dụng để tìm ra mối quan hệ của một tập dữ liệu thông qua một thiết kế kiến trúc chứa nhiều tầng ẩn (hidden layer), mỗi tầng lại chứa nhiều nơ ron. Các nơ ron được kết nối với nhau và độ mạnh yếu của các liên kết được biểu hiện qua trọng số liên kết.

Lập trình thông thường có thể làm được rất nhiều phần mềm lớn, như tính toán mô phỏng các vụ nổ hạt nhân trong siêu máy tính ở các phòng thí nghiệm, hoặc tái hiện các tế bào ở cấp độ phân tử để phân tích các thử nghiệm thuốc. Một siêu máy tính có thể tính toán được nhiều tỉ phép tính trên giây, tuy nhiên lập trình thông thường lại gặp khó khăn trong việc nhận ra các mẫu đơn giản, ví dụ như nhận diện mặt người, điều mà một bộ não sinh học xử lý nhanh và chính xác hơn nhiều.

Áp dụng với các kỹ thuật học sâu, mạng nơ ron nhân tạo hiện nay đang được áp dụng để giải quyết những vấn đề mà lập trình theo logic thông thường khó có thể giải quyết được. Do đó, mạng nơ ron nhân tạo đang nhanh chóng trở nên phổ biến, và là xu thế trên nhiều lĩnh vực.

**I.2.2. Một số mạng nơ ron nhân tạo**

Có hai kiểu mạng nơ ron chính: mạng nơ ron truyền thằng (feedforward neural network) và mạng nơ ron hồi quy (recurrent neural network).

Mạng truyền thẳng và hồi quy được minh họa như sau:

**A picture containing diagram

Description automatically generated**

**Hình ảnh mạng nơ ron truyền thẳng**

Ở mạng nơ ron truyền thẳng, các nút ở tầng thứ n+1 kết nối với tất cả các nút ở tầng thứ n, do mạng nơ ron có nhiều tầng ẩn nên ta thấy mạng nơ ron sẽ kéo dài trong không gian và không có vòng tuần hoàn nào. Mạng nơ ron truyền thẳng là mạng nơ ron được sử dụng nhiều nhất hiện nay.

Shape

Description automatically generated with medium confidence

**Hình ảnh mạng nơ ron hồi qui**

Mỗi nơ-ron tại một tầng của mạng nơ-ron hồi qui (trừ tầng đầu) đều nhận vào hai loại đầu vào: dữ liệu nhập và đầu ra của tầng trước. Mạng nơ ron hồi quy có ít nhất một đường dẫn tuần hoàn. Vì có một đường dẫn tuần hoàn, nên mạng nơ ron hồi quy có thể gây ra vòng lặp vô cực. Tuy nhiên, mạng nơ ron tuần hoàn có một ứng dụng quan trọng là chúng có thể nhận diện cho các giai đoạn thời gian khác nhau

**I.2.3. Thuật toán lan truyền ngược**

Phương pháp phổ biến nhất để tối ưu mạng neuron đa tầng chính là gradient descent (GD). Để áp dụng GD, chúng ta cần tính được gradient của hàm mất mát theo từng ma trận trọng số W(l) và vector điều chỉnh b(l) với l là chỉ số của 1 tầng trong mạng nơ ron.

Giả sử J(W, b, X, Y) là một hàm mất mát của bài toán, trong đó W, b là tập hợp tất cả các ma trận trọng số và vector điều chỉnh. X, Y là cặp dữ liệu huấn luyện với mỗi cột tương ứng với một điểm dữ liệu. Để có thể áp dụng các phương pháp gradient descent, chúng ta cần tính được các ∇W(l)J; ∇b(l)J, ∀l = 1, 2, . . . , L.

Sử dụng hàm mất mát là hàm sai số trung bình bình phương:

J(W, b, X, Y)== với là đầu ra thực tế, là đầu ra của mạng nơ ron.

Đạo hàm riêng của hàm mất mát theo chỉ một thành phần của ma trận trọng số của tầng đầu ra:

trong đó và vì +

Tương tự ta có đạo hàm của hàm mất mát theo hệ số tự do :

Với các đạo hàm tại các tầng l<L ta có đạo hàm tương tự bằng quy nạp:

với =

=

=

Tương tự ta có đạo hàm tại các tầng l<L theo hệ số tự do

Nhận thấy rằng trong những công thức trên, việc tính các đóng một vài trò quan trọng. Hơn nữa, để tính được giá trị này, ta cần tính được các . Nói cách khác, ta cần tính ngược các giá trị này từ tầng cuối cùng. Vì vậy tên gọi lan truyền ngược xuất phát từ đây.

Tóm tắt quá trình lan truyền ngược:

1. Lan truyền thuận: Với 1 giá trị đầu vào x, tính giá trị đầu ra của mạng, trong quá trình tính toán, lưu lại các giá trị tại mỗi tầng.
2. Với mỗi nút j ở tầng đầu ra, tính:

1. Với l=L-1, L-2…,1 tính:

=

1. Cập nhật gradient cho từng thành phần:

**I.3. Mạng nơ ron tích chập**

**I.3.1 Giới thiệu mạng nơ ron tích chập**

Mạng nơ ron tích chập (CNN) là một trong những mạng truyền thẳng đặc biệt. Mạng nơ ron tích chập là một mô hình học sâu phổ biến và tiên tiến nhất hiện nay. Hầu hết các hệ thống nhận diện và xử lý ảnh hiện nay đều sử dụng mạng nơ ron tích chập vì tốc độ xử lý nhanh và độ chính xác cao. Trong mạng nơ ron truyền thống, các tầng được coi là một chiều, thì trong mạng nơ ron tích chập, các tầng được coi là 3 chiều, gồm: chiều cao, chiều rộng và chiều sâu. Mạng nơ ron tích chập có hai khái niệm quan trọng: kết nối cục bộ và chia sẻ tham số. Những khái niệm này góp phần giảm số lượng trọng số cần được huấn luyện, do đó tăng nhanh được tốc độ tính toán.

**I.3.2 Mô hình mạng nơ ron tích chập**

Có ba tầng chính để xây dựng kiến trúc cho một mạng nơ ron tích chập:

1. Tầng tích chập (convolutional layer)
2. Tầng gộp (pooling layer)
3. Tầng kết nối đầy đủ (fully connected layer)

Tầng kết nối đầy đủ giống như các mạng nơ ron thông thường, và tầng chập thực hiện tích chập nhiều lần trên tầng trước. Tầng gộp có thể làm giảm kích thước mẫu. Ở các mạng nơ ron tích chập, kiến trúc mạng thường chồng ba tầng này để xây dựng kiến trúc đầy đủ.

Diagram

Description automatically generated

**Hình ảnh cấu trúc mạng nơ ron tích chập**

**I.3.3 Xây dựng mạng nơ ron tích chập**

**I.3.3.1 Tích chập**

Là một cửa sổ trượt (Sliding Windows) trên một ma trận như mô tả hình dưới:

**A picture containing shape

Description automatically generated**

Sliding Window hay còn gọi là kernel, filter hoặc feature detect là một ma trận có kích thước nhỏ như trong ví dụ trên là 3×3.

Stride: đối với phép tích chập hoặc phép pooling, độ trượt *S* ký hiệu số pixel mà cửa sổ sẽ di chuyển sau mỗi lần thực hiện phép tính.

Zero-padding: Zero-padding là tên gọi của quá trình thêm *P* số không vào các biên của đầu vào. Giá trị này có thể được lựa chọn thủ công hoặc một cách tự động.

Convolution hay tích chập là nhân từng phần tử bên trong ma trận 3×3 với ma trận bên trái. Kết quả được một ma trận gọi là Convoled feature được sinh ra từ việc nhận ma trận Filter với ma trận ảnh 5×5 bên trái. Kích thước của ma trận kết quả được tính theo công thức O= với O là kích thước 1 chiều của ma trận kết quả (chiều dài hoặc chiều rộng), I là kích thước chiều tương ứng của ma trận đầu vào, K là kích thước chiều tương ứng của kernel, P là số lượng zero-padding, S là bước nhảy (stride).

**I.3.3.2 Kết nối cục bộ**

Trong xử lý hình ảnh, thông tin của hình ảnh là các điểm ảnh (pixel). Nếu sử dụng mạng được kết nối đầy đủ, chúng ta sẽ có rất nhiều tham số. Ví dụ, một hình ảnh RGB có kích thước 512x512 pixel sẽ có 786432 (= 512 x 512 x 3) tham số đầu vào. Vì vậy, nếu chúng ta sử dụng kiến trúc mạng nơ ron trong hình sau:

**Diagram

Description automatically generated**

Hình trên cho thấy nếu áp dụng mạng nơ ron được kết nối đầy đủ, toàn bộ kiến trúc mạng sẽ cần tính toán hơn 3 triệu nơ ron. Số lượng lớn các nơ ron làm cho toàn bộ quá trình học rất chậm và dẫn đến quá tải so với khả năng tính toán của máy tính hiện tại.

Qua một vài nghiên cứu về xử lý ảnh, các nhà nghiên cứu đã nhận thấy rằng các điểm ảnh thường có xu hướng kết nối với các điểm ảnh khác nằm gần nó hơn là những điểm ảnh ở xa. Vì vậy, có thể chuyển đổi từ kiến trúc mạng kết nối đầy đủ sang kiến trúc mạng kết nối cục bộ nhằm giảm tài nguyên tính toán. Đây là ý tưởng chính của CNN.

Text

Description automatically generated with medium confidence

**Hình ảnh mô phỏng kết nối cục bộ**

**I.3.3.3 Chia sẻ tham số**

Chức năng chính của lớp này là hỗ trợ làm giảm tối đa số lượng những tham số trong mạng CNN. Vì trong mỗi Convolution sẽ bao gồm các Feature Map khác nhau, mỗi Feature Map lại giúp Detect một vài Feature trong ảnh.

**I.3.3.4 Tầng gộp**

Tầng gộp sẽ giảm kích cỡ của hình ảnh ngay sau khi thực hiện tích chập, giúp giữ lại các đặc điểm, tính chất nổi bật nhất của ảnh. Điều này cho phép giảm mức độ tính toán khi hình ảnh có kích thước quá lớn, đồng thời không làm mất đi các đặc điểm quan trọng của ảnh.

Tuy đã sử dụng các mạng được kết nối cục bộ và chia sẻ tham số, tuy nhiên số lượng tham số trong mang nơ ron vẫn là quá lớn. So với một tập dữ liệu tương đối nhỏ, nó có thể gây ra tình trạng quá khớp (overfitting). Vì vậy, mạng nơ ron nhân tạo thường chèn các tầng gộp vào mạng.

**A picture containing diagram

Description automatically generated**

**Hình ảnh minh họa tầng max pooling**

**I.3.3.5 Tầng kết nối đầy đủ**

Tầng này giống như mạng nơ ron truyền thống: các nơ ron ở tầng trước sẽ kết nối với một nơ ron ở tầng tiếp theo và tầng cuối cùng là đầu ra. Để có thể đưa hình ảnh từ những tầng trước vào, ta phải dàn phẳng dữ liệu ra thành một véc tơ nhiều chiều. Cuối cùng là sử dụng hàm softmax để thực hiện phân loại đối tượng.

**A picture containing chart

Description automatically generated**

**Hình ảnh minh họa lớp kết nối đầy đủ**

**I.3.3.6 Mô hình quá khớp**

Cấu trúc một mạng nơ ron tích chập thường rất lớn, trong mạng có rất nhiều nơ ron, kết nối, trọng số cần huấn luyện. Tuy nhiên dữ liệu huấn luyện thường không đủ lớn để huấn luyện hoàn thiện một mạng nơ ron lớn. Do đó xảy ra hiện tượng quá khớp, đó là mạng khi huấn luyện cho hiệu quả rất cao nhưng khi đưa vào sử dụng trong thực tế thì hiệu quả lại kém.

Chart

Description automatically generated

**Hình ảnh minh họa quá khớp**

Hình ảnh trên cho thấy khi huấn luyện mô hình kết quả thu được là đường màu xanh thì hiện tượng quá khớp đã xảy ra, kết quả là đường màu đen là kết quả tốt.

Một trong những phương pháp tránh quá khớp đó là là giảm trọng số trong lúc huấn luyện. Dropout là một trong những kỹ thuật nổi tiếng và khá phổ biển để khắc phục vấn đề này. Dropout đặt đầu ra của mỗi nơ ron ẩn thành 0 với xác suất 0,5. Vì vậy, các nơ ron này sẽ không đóng góp vào lan truyền tiến, do đó và sẽ không tham gia vào lan truyền ngược. Thông thường, đối với các đầu vào khác nhau, mạng nơ ron xử lý dropout theo một cấu trúc khác nhau.

Một cách khác để cải thiện việc việc quá khớp là tăng lượng dữ liệu. Chúng ta có thể phản chiếu hình ảnh, lộn ngược hình ảnh, lấy mẫu hình ảnh,… Những cách này sẽ tăng số lượng dữ liệu huấn luyện. Vì vậy, nó có khả năng ngăn chặn quá khớp.

**I.3.3.7 Batch normalization**

**Batch Normalization** là một phương pháp hiệu quả khi training một mô hình mạng nơ ron. Mục tiêu của phương pháp này chính là việc muốn chuẩn hóa các feature (đầu ra của mỗi layer sau khi đi qua các activation) về trạng thái **zero-mean** với độ lệch chuẩn 1.

Text

Description automatically generated with medium confidence

**Hình ảnh minh họa batch normalization**

Batch normalization có thể giúp chúng ta tránh được hiện tượng giá trị của x rơi vào khoảng bão hòa sau khi đi qua các hàm kích hoạt phi tuyển. Vậy nên nó đảm bảo rằng không có sự kích hoạt nào bị vượt quá cao hoặc quá thấp. Điều này giúp cho các weights mà khi không dùng BN có thể sẽ không bao giờ được học thì nay lại được học bình thường. Điều này giúp chúng ta làm giảm đi sự phụ thuộc vào giá trị khởi tạo của các tham số.

Batch Normalization còn có vai trò giúp cho việc giảm thiểu overfiting. Sử dụng batch normalization, chúng ta sẽ không cần phải sử dụng quá nhiều dropput và điều này rất có ý nghĩa vì chúng ta sẽ không cần phải lo lắng vì bị mất quá nhiều thông tin khi dropout weigths của mạng.

**I.3.3.8 Giới thiệu mạng VGG16**

VGG16 là một Kiến trúc mạng nơ-ron liên kết (CNN) đơn giản và được sử dụng rộng rãi được sử dụng cho ImageNet, một dự án cơ sở dữ liệu trực quan lớn được sử dụng trong nghiên cứu phần mềm nhận dạng đối tượng trực quan. Kiến trúc VGG16 được phát triển và giới thiệu bởi Karen Simonyan và Andrew Zisserman từ Đại học Oxford, vào năm 2014, thông qua bài báo của họ “Các mạng kết hợp rất sâu để nhận dạng hình ảnh quy mô lớn”. 'VGG' là tên viết tắt của Visual Geometry Group, là một nhóm các nhà nghiên cứu tại Đại học Oxford, những người đã phát triển kiến ​​trúc này, và '16' ngụ ý rằng kiến ​​trúc này có 16 lớp

Mô hình VGG16 đạt độ chính xác trong bài kiểm tra top 5 là 92,7% trong ImageNet, đây là tập dữ liệu của hơn 14 triệu hình ảnh thuộc 1000 lớp. Đây là một trong những mô hình nổi tiếng được gửi đến Thử thách nhận dạng hình ảnh quy mô lớn ImageNet (ILSVRC) vào năm 2014. Nó đã thực hiện các cải tiến so với kiến ​​trúc AlexNet bằng cách thay thế các bộ lọc có kích thước hạt nhân lớn (11 và 5 trong lớp chập đầu tiên và thứ hai, tương ứng) với nhiều bộ lọc kích thước hạt nhân 3x3 lần lượt.

Kiến trúc mạng VGG16 gồm 16 lớp với 13 lớp tích chập và 3 lớp kết nối đầy đủ. Mạng sử dụng ảnh đầu vào có kích thước 224x224x3, kernel có kích thước 3x3, lớp gộp sử dụng là max pooling với pooling size có kích thước 2x2 và stride = (2x2).

Diagram

Description automatically generated

**Hình ảnh minh họa mạng VGG16**

**I.4 Kết luận**

Mạng nơ ron nhân tạo là một chuỗi các thuật toán được sử dụng để tìm ra mối quan hệ của một tập dữ liệu thông qua cơ chế vận hành của bộ não sinh học. Mạng nơ ron nhân tạo thường được huấn luyện qua một tập dữ liệu chuẩn cho trước, từ đó có thể đúc rút được kiến thức từ tập dữ liệu huấn luyện, và áp dụng với các tập dữ liệu khác với độ chính xác cao. Các phương pháp sử dụng để huấn luyện mạng nơ ron nhân tạo ngày càng tối ưu hơn về mặt tính toán và phục vụ cho nhiều mục đích khác nhau.

Hiện nay, kiến trúc mạng nơ ron ngày càng được hoàn thiện cho nhiều nhiệm vụ, trong đó mạng nơ ron tích chập được chú ý rất nhiều vì tính hiệu quả trong thị giác máy tính. Mạng nơ ron tích chập với các cải tiến góp phần giảm thời gian tính toán và tăng độ chính xác hứa hẹn sẽ là một trong những phương pháp được áp dụng rất nhiều vào thực tế trong tương lai.

**CHƯƠNG II.**

**Bài toán phân loại hình ảnh bằng mạng nơ ron tích chập**

**II.1. Bài toán phân loại hình ảnh**

Phân loại hình ảnh là một nhiệm vụ phổ biến của thị giác máy tính. Nhiệm vụ của bài toán là từ một ảnh đầu vào (input) cho ra một nhãn của hình ảnh (output).

Thực tế, thị giác góp phần tạo nên 80-85% nhận thức của con người về thế giới. Hàng ngày, mỗi người phải thực hiện phân loại trên bất kỳ dữ liệu hình ảnh nào mà chúng ta bắt gặp.

Do đó, mô phỏng nhiệm vụ phân loại với sự trợ giúp của mạng nơ-ron là một trong những ứng dụng đầu tiên của thị giác máy tính mà các nhà nghiên cứu nghĩ đến.

**Cách thức hoạt động**

Máy tính xử lý một hình ảnh dưới dạng pixel. Theo đó, hình ảnh chỉ là một mảng ma trận, và kích thước của ma trận phụ thuộc vào độ phân giải hình ảnh.

Do đó, xử lý hình ảnh là tiến hành phân tích dữ liệu toán học với sự trợ giúp của các thuật toán. Các thuật toán này chia nhỏ hình ảnh thành một tập hợp các đặc điểm nổi bật, giúp giảm khối lượng công việc của bộ phân loại cuối cùng.

Quá trình trích xuất đặc điểm là bước quan trọng nhất trong việc phân loại hình ảnh. Quá trình phân loại phụ thuộc phần lớn vào dữ liệu được cung cấp cho thuật toán. Một bộ dữ liệu phân loại tốt phải đảm bảo các yêu cầu về sự cân bằng của dữ liệu, chất lượng của ảnh và chú giải kèm theo.

Quá trình phân loại hình ảnh diễn ra như sau:

1. Bước 1: từ ảnh đầu vào, với các trọng số mà mô hình đã được học, mô hình trích xuất các đặc trưng của ảnh.
2. Bước 2: từ các đặc trưng được trích xuất, chúng được qua một vài lớp kết nối đầy đủ và cuối cùng đi qua hàm softmax để phân loại. Vị trí có trọng số (xác suất) lớn nhất chính là nhãn được dự đoán.

Bản chất của quá trình đào tạo mô hình là từ các dữ liệu đầu vào sẽ thực hiện quá trình lan truyền thuận, sau đó lan truyền ngược để cập nhật gradient cho tất cả các nơ ron trong mô hình. Quá trình dự đoán nhãn hình ảnh thì chỉ xảy ra quá trình lan truyền thuận.

**II.2. Thực nghiệm phân loại hình ảnh bằng mạng nơ ron tích chập**

**II.2.1. Mô hình và tập dữ liệu**

Mô hình sử dụng trong bài toán phân loại hình ảnh là VGG16 với 13 lớp tích chập, 3 lớp kết nối đầy đủ. Đầu vào là ảnh RGB kích thước 224x224x3, đầu ra là một lớp kết nối đầy đủ với 30 lớp đầu ra và hàm kích hoạt là hàm softmax.

A picture containing letter

Description automatically generated

**Hình ảnh kiến trúc mạng VGG16 sử dụng trong thực nghiệm**

Tập dữ liệu sử dụng là tập dữ liệu ảnh hoa quả với 30 lớp khác nhau, trong đó chia làm 3 phần : train với 14934 ảnh, validation với 1491 ảnh và test với 1494 ảnh. Link tập dữ liệu <https://drive.google.com/drive/folders/1nUmTVa6fKf2vNvF0bywhUvMqmLINnqC3?usp=sharing>.

Graphical user interface, website

Description automatically generated

**Hình ảnh minh họa tập dữ liệu**

**II.2.2. Kết quả thu được**

Với mô hình và bộ dữ liệu trên, kết quả thu được khá tốt với độ chính xác trên tập đánh giá đạt 99,46% và loss = 0,1403, độ chính xác trên tập test đạt 99,46% và loss=0,1399 chỉ sau 20 epochs.

Graphical user interface

Description automatically generated

**Hình ảnh so sánh giữa tập huấn luyện và tập đánh giá**

Như vây, mô hình dự đoán khá chính xác trên tập kiểm tra, cho thấy mô hình tránh được hiện tượng quá khớp.

Graphical user interface

Description automatically generated

**Hình ảnh minh họa kết quả dự đoán của mô hình trên tập kiểm tra**

Đường link dẫn đến file mô hình đã được đào tạo (file batchProject1VGG16.h5) : <https://drive.google.com/file/d/1-1t7vOfqL756d-upvAG55fKuGEmGQC23/view?usp=sharing>

**II.2.3. Hướng phát triển trong tương lai**

Bài toán phân loại hình ảnh là bài toán phổ biến của thị giác máy tính. Vì vậy nó là bài toán con có mặt trong hầu hết các bài toán lớn hơn của thị giác máy tính như nhận diện vật thể, nhận diện khuôn mặt,...Vì vậy trong tương lai, mô hình có thể phát triển thành mô hình nhận diện trái cây để phát hiện và loại trái cây khác loại ứng dụng trong công nghiệp như sản xuất rượu trái cây, trái cây xấy khô,…Ứng dụng trong hệ thống theo dõi và chăm sóc cây nông nghiệp tự động.

**Kết luận**

**Kết quả đã thực hiện được của project**

Ngày nay, với sự phát triển của khoa học và công nghê, con người liên tục tạo ra những hệ thống thông minh nhằm giải phóng sức lao động của con người, cải thiện những hệ thống truyền thống. Cùng với mục tiêu đó, bài toán phân loại hình ảnh góp phần là một bài toán nhỏ trong một hệ thống như vậy. Đề tài ‘Tìm hiểu mạng nơ ron tích chập và ứng dụng mạng nơ ron tích chập trong bài toán phân loại hình ảnh’ đã đạt được những mục tiêu sau :

1. **Chương 1 :** Báo cáo đã trình bày tổng quan về lịch sử hình thành, cũng như các kiến thức về nơ ron nhân tạo, mạng nơ ron nhân tạo và mạng nơ ron tích chập. Về mạng nơ ron, báo cáo đã làm rõ được cấu trúc của một nơ ron nhân tạo, cách hoạt động của từng thành phần nơ ron. Ngoài ra, báo cáo cũng đã làm rõ cách mạng nơ ron nhân tạo thực hiện quá trình huấn luyện bằng hình thức lan truyền ngược. Về các kiến trúc mạng, báo cáo đã giới thiệu sơ qua về hai kiến trúc mạng phổ biến là mạng nơ ron truyền thẳng và mạng nơ ron hồi quy. Trong đó có đi sâu vào phân tích một mạng nơ ron truyền thẳng đặc biệt là mạng nơ ron tích chập.
2. **Chương 2 :** Báo cáo đã trình bày ý tưởng của bài toán phân loại hình ảnh, mô tả mô hình và tập dữ liệu sử dụng trong quá trình huấn luyện mô hình. Báo cáo cũng mô phỏng kết quả của quá trình huấn luyện cũng như mô phỏng khả năng dự đoán của mô hình.

Sau khi hoàn thành Project 1, em đã hiểu được cấu trúc, cách học và đưa ra dự đoán của một mạng nơ ron, các kỹ thuật cải thiện độ chính xác cũng như tăng nhanh quá trình hội tụ trong đào tạo mạng nơ ron. Từ những kiến thức đạt được này, em sẽ áp dụng và phát triển vào các project tiếp theo.

**TÀI LIỆU THAM KHẢO**

|  |  |
| --- | --- |
| [1] | N.T.Tuấn/ Blog Deep learning. URL: https://nttuan8.com/gioi-thieu-ve-deep-learning/ |
| [2] | N.T.Tuấn/ Book Deep learning cơ bản. URL: https://nttuan8.com/sach-deep-learning-co-ban/ |
| [3] | V.K.Tiệp/ Book Machine learning cơ bản. URL: https://github.com/tiepvupsu/ebookMLCB/blob/master/book\_ML.pdf |