

Understand and Build CNN from the Grounth Up and Intuition

Nhóm AIO_TimeSeries

Ngày 18 tháng 11 năm 2025

Bài blog này tiếp cận CNN từ góc nhìn của Computer Vision và giải thích từ căn bản để người đọc có thể hiểu được ý nghĩa thực sự đằng sau CNN là gì. Blog sẽ bao quát các chủ đề sau:

Phần 1: Vấn đề với MLP trong Phân loại Ảnh Phần 2: Tổng quan kiến trúc CNN
Từ Pixels đến Output qua các lớp Conv, ReLU, Pooling

Phần 3: CNN Operation
Convolution, Padding, Stride, ReLU trong CNN

Phần 4: CNN Backpropagation
Cơ chế lan truyền ngược trong mạng tích chập

Phần 5: Advance CNN
Pooling, Batch Normalization, 1x1 Convolution

Phần 6: Thực hành Code CNN dưới góc nhìn lập trình viên
Triển khai CNN bằng tư duy lập trình hướng đối tượng OOP

Phần 1: Giới thiệu về Computer Vision & Tensor

Trong phần này, chúng ta tìm hiểu cách hình ảnh được biểu diễn dưới dạng các Tensor 1D, 2D và 3D. [\[5\]](#) **Vấn đề của MLP (Multi-Layer Perceptron):** Khi xử lý hình ảnh lớn, số lượng node đầu vào tăng vọt khiến MLP chạy rất chậm và dễ bị quá tải tham số. Giải pháp là sử dụng lớp Convolution để đơn giản hóa các đặc trưng (features) trước khi đưa vào MLP. CNN không thay thế MLP mà đóng vai trò là bộ trích xuất đặc trưng (feature extractor) mạnh mẽ hơn.

Phần 2: Tổng quan kiến trúc CNN

Kiến trúc CNN bao gồm các thành phần chính giúp xử lý thông tin không gian hiệu quả. Quy trình tính toán tổng quát:

$(\text{Pixels} + \text{bias}) \rightarrow (\text{Conv} \rightarrow \text{ReLU}) \times 2 \rightarrow \text{Pooling} \rightarrow (\text{Flatten} + \text{bias}) \rightarrow \text{Softmax} \rightarrow \text{Output}$

Sử dụng công cụ trực quan hóa như CNN Explainer để thấy sự khác biệt trước và sau khi qua các lớp.

Phần 3: Đi sâu vào các phép toán CNN

Dựa trên tài liệu của Stanford CS231n:

- **Convolution:** Phép tính tích chập lặp lại để trích xuất đặc trưng cục bộ.
- **Padding:** Khi nào cần thiết để giữ lại kích thước của matrix?
- **Stride:** Bước nhảy của kernel. $\text{Stride} = 1$ nghĩa là kernel di chuyển từng pixel. Ảnh hưởng của Stride đến việc trích xuất đặc trưng cục bộ (local) và toàn cục (global).
- **ReLU:** Đưa tính phi tuyến vào mô hình và loại bỏ nhiễu từ các giá trị âm.

Phần 4: CNN Backpropagation

Giải thích cách cập nhật trọng số cho các bộ lọc (filters) thông qua đạo hàm của phép toán tích chập.

Phần 5: CNN Nâng cao & Lập trình OOP

Tìm hiểu về Pooling (Max/Average), Batch Normalization và đặc biệt là ý nghĩa của phép tích chập 1×1 . Cuối cùng là triển khai code thực tế bằng Python theo hướng đối tượng (OOP).

References

- [1] In Depth. *How CNN work, in depth*. 2024. URL: https://youtu.be/JB8T_zN7ZC0.
- [2] Quora. *What are the main role of ReLU Activation Function in deep learning*. 2024. URL: <https://www.quora.com/What-are-the-main-role-of-ReLU-Activation-Function-in-deep-learning>.
- [3] Brandon Rohrer. *How Convolutional Neural Networks work*. 2024. URL: https://brandonrohrer.com/how_convolutional_neural_networks_work.html.
- [4] Nguyễn Tất Tuân. *Bài 6: Convolutional Neural Network*. 2024. URL: <https://nttuan8.com/bai-6-convolutional-neural-network/>.
- [5] Stanford University. *CS231n: Convolutional Neural Networks for Visual Recognition*. 2024. URL: <https://cs231n.github.io/convolutional-networks/>.