Convolutional Neural Networks

AUTHORS: Jed Pulley & Keshav Sharan Pachipala

DO NOT POLLUTE! AVOID PRINTING, OR PRINT 2-SIDED MULTIPAGE.

1   Introduction

In recent years, Convolutional Neural Networks (CNNs) have emerged as a cornerstone technology in the field of deep learning, particularly in the domain of computer vision. CNNs are adept at extracting intricate patterns and features from spatially structured data. Some fields that benefit from CNNs include:

1. Image Classification

2. Object Detection

3. Medical Image Analysis

In this paper, we won't go into the exact specifics of what a convolution is since it's slightly different in the discrete and continuous cases, as well as slightly different in a CNN context, but if you're really curious about it, 3Blue1Brown has some excellent videos visually explaining the mathematical concepts behind a convolution in his videos linked below [1] and [2].

2   Understanding the Architecture

2.1   NNs vs CNNs

CNNs and Neural Networks are similar in many ways, but also have drastically different architectures. From a high level, the classic neural network is shown as an input layer of neurons, a hidden layer, and an output layer, shown below:

<center>input layer   hidden layer 1   hidden layer 2   output layer</center>
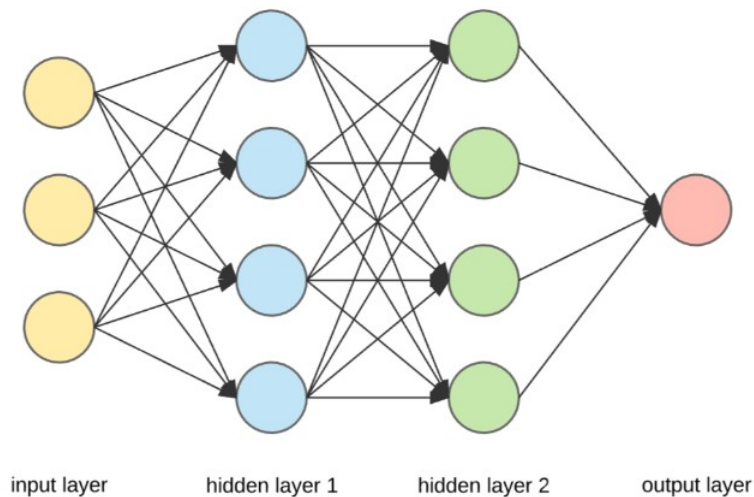
However, the architecture of a convolution neural network looks much different. There isn't one right way to structure it, but typically a CNN will have the input, a few convolution and pooling layers, a flattening step, and then finally a fully connected layer that leads to an output, as shown below:
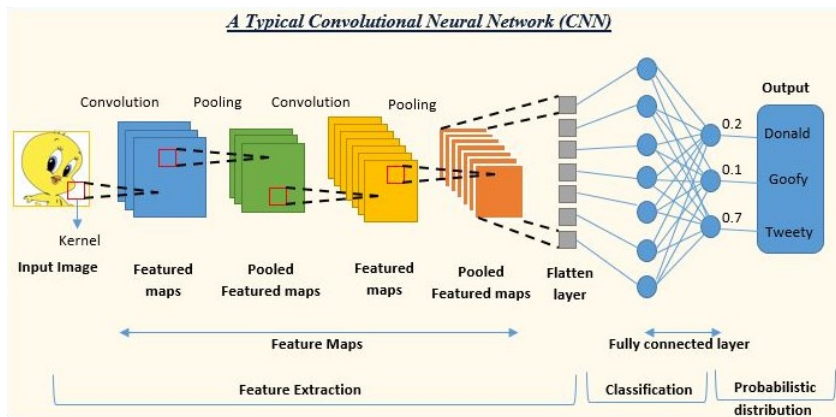
How you structure your CNN is an area of active research today. However, the above is a completely valid example.

2.2 Kernels

Fundamental to the idea of convolutions are kernels. In the 2D image processing case, a kernel often takes the form of a 3x3 grid of numbers that are "convolved" with the image. This convolution step means sliding the kernel across the image pixel by pixel, multiplying each element with it's overlapping counterpart, summing them up, and then dividing them by the kernel size (3x3 having a size of 9). The output is a single pixel in the output image/matrix. You scroll through the entire image performing the same operation until complete. Since the operative pixel is the center of the kernel, convolution runs into the issue of overlapping the image. So to avoid this, with a 3x3 kernel for example, a single pixel of padding is added to the image so that the kernel will fit in the first row/column.
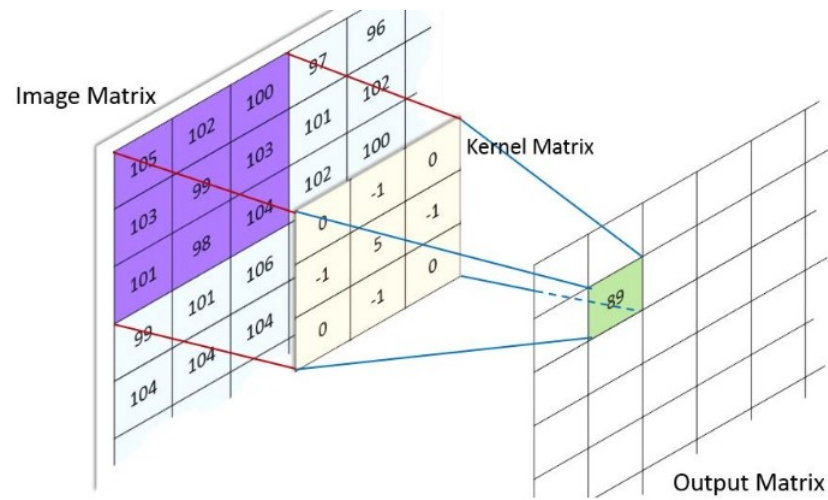
2.3 Layers

(a) Convolution Layer The convolution layer is the most fundamental aspect of a CNN.

(b) Pooling Layer

(c) Flattening Layer

(d) Fully Connected Layer

(e) Output Layer

2.4 Feature Hierarchy

3 Advanced CNN Architectures

(a) VGGNet

(b) ResNet

(c) Inception Network

4 Future Directions and Trends

(a) TO-DO

(b) TO-DO

(c) TO-DO

5 Resources

[**1**] 3Blue1Brown - But What is a Convolution?
`https://www.youtube.com/watch?v=KuXjwB4LzSA&t=1s`

[**2**] 3Blue1Brown - Convolutions — Why X+Y in probability is a beautiful mess
`https://www.youtube.com/watch?v=IaSGqQa5O-M`

[**3**] far1din - Convolutional Neural Networks from Scratch — In Depth
`https://www.youtube.com/watch?v=jDe5BAsT2-Y`

[**4**] Younesi, A., Ansari, M., Fazli, M. A., Ejlali, A., Shafique, M., & Henkel, J. (2024). A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends. arXiv preprint arXiv:2402.15490
`https://arxiv.org/abs/2402.15490`