

真-极简神经网络入门 CNN篇

作者: 骰子AI



大纲

-
- 代码地址: [rexrex9/basic_neural_networks_pytorch](https://github.com/rexrex9/basic_neural_networks_pytorch): 最入门的神经网络示例代码,pytorch版 (github.com) 其中的chapter2_cnn
1. 卷积神经网络概述
 2. 卷积层
 3. 填充
 4. 步幅
 5. 通道
 6. 池化层
 7. 实战: MNIST手写数字识别

1. 卷积神经网络概述

- 卷积神经网络(Convolutional Neural Networks, CNN) 从结构上可分为卷积层，池化层，激活层。
 1. 池化一般指取输入值的平均或最大值等操作，只是在卷积神经网络中是对二维张量的池化。
 2. 激活指的就是Sigmoid, Tanh这样的非线性激活函数。
 3. CNN的重点在于卷积层。
 - 卷积层具备3种超参，分别是填充，步幅，通道。

1. 卷积神经网络概述



- 基础神经网络对于图片信息的局限性：
 1. 假设图的分辨率为 960×600 , 忽略通道数的情况下。如果用基础的MLP神经网络进行消息传递。则模型需要训练的参数至少需要 960×600 个。
 2. 基础线性层并无法学到上下左右的空间信息。
 3. 像素一般只和周围的像素产生信息影响。每次传递时计算全部像素是一种浪费。
- 卷积的优势：
 1. 减少训练参数。
 2. 学到空间信息。
 3. 学到局部信息。

2. 卷积层

1	2	3
4	5	6
7	8	9

*

0	1
2	3

=

25	31
43	49

1	2
4	5

*

0	1
2	3

=

25

$$1 \times 0 + 2 \times 1 + 4 \times 2 + 5 \times 3 = 25$$

$$2 \times 0 + 3 \times 1 + 5 \times 2 + 6 \times 3 = 31$$

$$4 \times 0 + 5 \times 1 + 7 \times 2 + 8 \times 3 = 43$$

$$5 \times 0 + 6 \times 1 + 8 \times 2 + 9 \times 3 = 49$$

- 卷积层指通过一个矩阵即卷积核 K (kernel) 对输入的矩阵 X 进行计算。默认情况下, 若输入的矩阵形状是 $[a, b]$, 卷积核的形状是 $[h, w]$ 。则输出矩阵 Y 的形状为 $[a - h + 1, b - w + 1]$
- 卷积核是模型需训练的参数。
- 输出矩阵 Y 中每个元素的计算公式为:

$$Y_{ij} = \text{sum}(X_{[i:i+h, j:j+w]} \odot K)$$

3. 填充

0	0	0	0	0
0	1	2	3	0
0	4	5	6	0
0	7	8	9	0
0	0	0	0	0

*

0	1
2	3

=

3	8	13	6
13	25	31	12
25	43	49	18
7	8	9	0

0	0
0	1

*

0	1
2	3

=

3

- 填充(Padding) 指的是将输入矩阵上下左右方向增加若干行或列，并用默认为0的数字填充以便调整形状。

4. 步幅

步幅为2

0	0	0	0	0
0	1	2	3	0
0	4	5	6	0
0	7	8	9	0
0	0	0	0	0

*

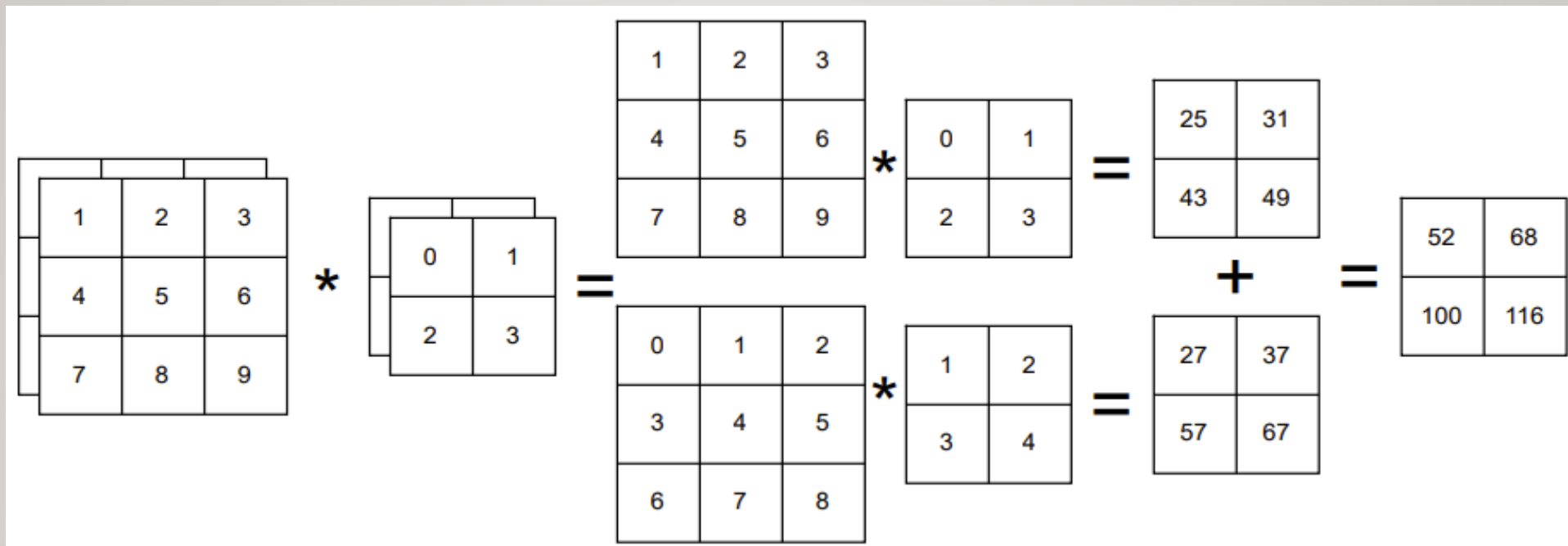
0	1
2	3

=

3	13
25	49

- 步幅(**Stride**) 指的是窗口滑动的距离。
- 步幅可以分为垂直步幅与水平步幅。

5. 通道



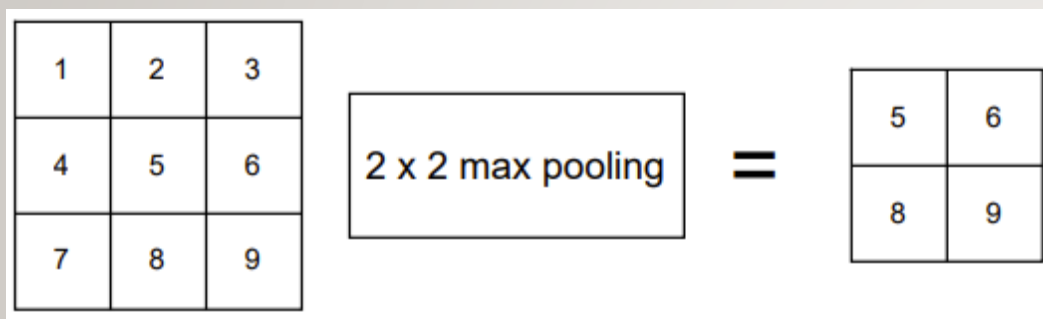
仅考虑输入通道

- 输入X: $[C_i, X_h, X_w]$
- 卷积核K: $[C_i, K_h, K_w]$
- 输入Y: $[Y_h, Y_w]$

多个输出通道

- 输入X: $[C_i, X_h, X_w]$
- 卷积核K: $[C_o, C_i, K_h, K_w]$
- 输入Y: $[C_o, Y_h, Y_w]$

6. 池化层



$$\text{Max} \left(\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 4 & 5 \\ \hline \end{array} \right) = \boxed{5}$$

$$\text{Avg} \left(\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 4 & 5 \\ \hline \end{array} \right) = \boxed{3}$$

- 池化(Pooling)通常分为最大值池化(Max Pooling)与平均值池化(Avg Pooling)。
- 最大值池化(Max Pooling): 取出窗口内的最大值。
- 平均值池化(Avg Pooling): 计算窗口内的平均值。
- 池化层同样具备填充与步幅的超参, 但不没有通道数的概念。
- 池化层是一种计算操作, 并没有模型参数需要训练。

7. 实战：MNIST手写数字识别



- 本次训练的目的在于使得模型通过手写数字的图片识别具体数字。
- MNIST数据集包含60000个训练样本，10000个测试样本。
- 可以直接通过 `torchvision.datasets.MNIST` 下载。

结束

