

池化层的作用

池化层通过将特征图的尺寸缩小，有效的减少了网络中需要处理的数据量。这样做的好处有：

- 降低计算成本：后续卷积层和全连接层的计算量会大幅减少，网络训练和推理更快。
- 减少参数：由于特征图尺寸减少，连接到后续全连接层的神经元数量也相应的减少，从而降低了整个网络的参数数量，有助于防止过拟合。
- 增强平移不变性：池化层通过对局部进行**特征聚合**来实现这一点。如果一个物体在图像中轻微移动，只要它仍然处于同一个池化窗口内，池化层的输出基本上不会变化，从而提高模型的鲁棒性和泛化能力。
- 提取最显著的特征：
 - 最大池化，它选择局部区域最大值作为输出，通常被认为是一种特征选择的方法，因为它保留最显著、最强烈的特征，而丢弃不太重要的信息。这使得网络能够更有效的捕捉关键模型。
 - 平均池化，它计算局部区域中所有值的平均值作为输出。这种方法更像是对特征进行平滑和概括，它保留区域中所有信息的整体趋势，而不是突出特定的强信号。

参数说明

- 池化层输入 X 的形状一般为 $[N, C, H, W]$
- 池化方法：我们只考虑最大值池化和平均池化
- 池化窗口尺寸： K_h, K_w
- 滑动步长： S_h, S_w
- 填充尺寸： P_h, P_w

池化过程只对特征图操作，对于不同的批次和通道都同样处理即可。

计算输出的特征图的尺寸：

$$H_{out} = \text{floor} \left(\frac{H + 2 \cdot P_h - K_h}{S_h} + 1 \right)$$
$$W_{out} = \text{floor} \left(\frac{W + 2 \cdot P_w - K_w}{S_w} + 1 \right)$$

前向传播

最大池化

最大池化，是找出池化窗口中的最大值作为输出，为了能在反向传播时，正确设置梯度，需要将选中的最大值的索引记录下来，在反向传播时的，梯度应该赋予相应的位置上。

一般，我们会创建一个与输出特征图相同形状的整形张量，用于记录选择的最大值在当前输入特征图中的索引。这个索引会初始化为 -1， 同时将输出特征图初始化为负无穷大。

在滑动窗口中寻找最大值时，要忽略掉填充的位置，即使在滑动窗口中最大值位于填充位置，也不能选择，我们要选择输入特征图中的当前滑动窗口的最大值。并记录其在当前特征图中的索引。

平均池化

平均池化的处理方式与最大池化的不同至于，对滑动窗口内的所有数值累加，最后再除以池化窗口元素总数，求得平均值作为输出。

池化窗口元素需要考虑填充区域，比如一个 3×3 的滑动窗口，尽管可能包含一些填充数据，但池化窗口的元素总数依然是 9。

反向传播

反向传播的输入梯度的形状为 $[N, C, H_{out}, W_{out}]$ ，而反向传播输出的梯度形状与前向传播输入的形状一致，为 $[N, C, H, W]$ 。

最大池化

我们要利用到前向传播时记录的索引张量。对于输出特征图梯度中的某个元素，我们利用与它关联的索引值，将梯度累积到这个索引值指向的输出梯度的位置即可。

平均池化

我们需要根据输出特征图中的梯度元素索引计算出它的池化窗口元素总数（包括填充），然后将梯度除以池化窗口元素总数，求得平均梯度，再分别累加到输入梯度的池化窗口中的所有有效元素（填充区域

应当跳过)。