XLAV Summary

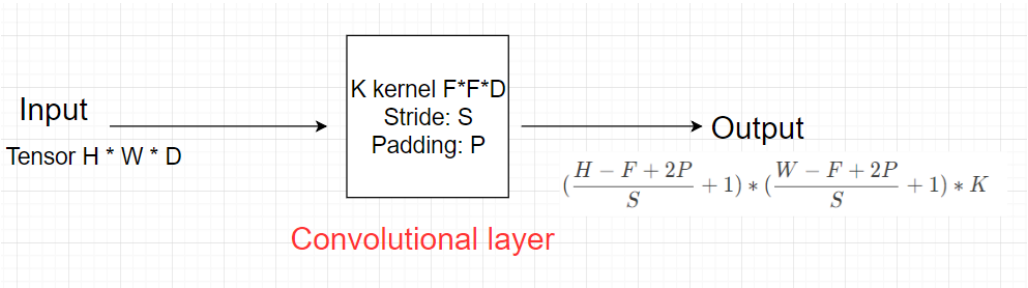
November 2023

1 Convolutional Neural Network

1.1 Convolutional layer tổng quát

Giả sử input của 1 convolutional layer tổng quát là tensor kích thước $H * W * D$. Kernel có kích thước $F * F * D$ (kernel luôn có depth bằng depth của input và F là số lẻ), stride: S, padding: P. Convolutional layer áp dụng K kernel. Output của layer là tensor 3 chiều có kích thước:

$$\left(\frac{H - F + 2P}{S} + 1\right) * \left(\frac{W - F + 2P}{S} + 1\right) * K$$



Còn khi đi qua lớp Pooling, công thức tính output sẽ như sau:

$$\lfloor\frac{H - F}{S} + 1\rfloor * \lfloor\frac{W - F}{S} + 1\rfloor * K$$

Ví dụ với mô hình LeNet, ta có code PyTorch:

```
import torch
from torch import nn
from d2l import torch as d2l

net = nn.Sequential(
    nn.Conv2d(1, 6, kernel_size=5, padding=2), nn.Sigmoid(),
    nn.AvgPool2d(kernel_size=2, stride=2),
    nn.Conv2d(6, 16, kernel_size=5), nn.Sigmoid(),
    nn.AvgPool2d(kernel_size=2, stride=2),
    nn.Flatten(),
    nn.Linear(16 * 5 * 5, 120), nn.Sigmoid(),
    nn.Linear(120, 84), nn.Sigmoid(),
    nn.Linear(84, 10))
```

và TensorFlow:

```
import tensorflow as tf

def net():
```

```
return tf.keras.models.Sequential([
    tf.keras.layers.Conv2D(filters=6, kernel_size=5, activation='sigmoid',
                           padding='same'),
    tf.keras.layers.AvgPool2D(pool_size=2, strides=2),
    tf.keras.layers.Conv2D(filters=16, kernel_size=5,
                           activation='sigmoid'),
    tf.keras.layers.AvgPool2D(pool_size=2, strides=2),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(120, activation='sigmoid'),
    tf.keras.layers.Dense(84, activation='sigmoid'),
    tf.keras.layers.Dense(10)])
```

Với ảnh đầu vào có kích thước là $28 * 28 * 1$, output cho từng layer của kiến trúc trên như sau:

- Lớp Conv2D thứ nhất: Do height bằng width, trong đó:
  - input size bằng 28 (chiều dài và chiều rộng của ảnh đầu vào là 28 pixel).
  - padding bằng 2
  - kernel size = 5
  - stride bằng 1 (mặc định nếu không truyền)

  nên ta có:
  $$\text{outputsize} = \frac{28 + 2 * 2 - 5}{1} + 1 = 28$$
  Do đó, output shape sau khi ảnh qua lớp Conv2D thứ nhất là $28 * 28 * 6$, với 6 là số lượng kênh.

- Lớp AvgPool thứ nhất:
  - input size là kích thước đầu vào sau khi đi qua lớp Conv2D thứ nhất $28 * 28$.
  - kernel size bằng 2
  - stride bằng 2
  $$\text{outputsize} = \lfloor \frac{28 - 2}{2} \rfloor + 1 = 14$$
  Do đó, output shape sau khi qua lớp AvgPool thứ nhất là $14 * 14 * 6$.

- Lớp Conv2D thứ hai:
  - input size là kích thước đầu ra của AvgPool thứ nhất $14 * 14$.
  - padding không được chỉ định nên mặc định bằng 0
  - kernel size bằng 5
  - stride bằng 1 (mặc định nếu không truyền)
  $$outputsize = \frac{14 + 2 * 0 - 5}{1} + 1 = 10$$
  Vì vậy, output shape sau khi qua lớp Conv2D thứ hai là $10 * 10 * 16$, với 16 là số lượng kênh.

- Lớp AvgPool thứ hai:
  - input size là kích thước của output từ lớp Conv2D thứ hai $10 * 10$.
  - kernel size bằng 2
  - stride bằng 2
  $$outputsize = \lfloor \frac{10 - 2}{2} \rfloor + 1 = 5$$
  Vậy output shape sau khi qua lớp AvgPool thứ hai là $5 * 5 * 16$, với 16 là số lượng kênh.