Pooling

Recap: Convolution

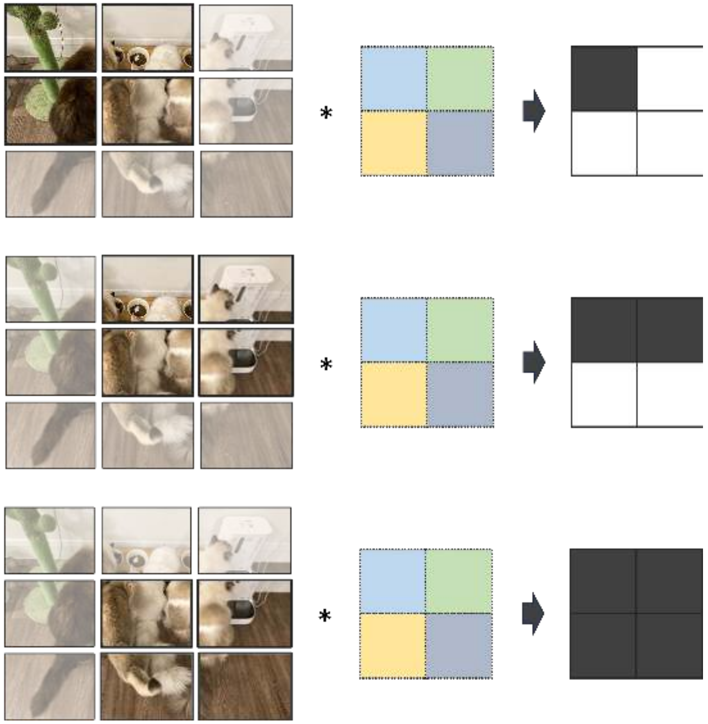
Input: $x \in \mathbb{R}^{C_1 \times H \times W}$

Output: $y \in \mathbb{R}^{C_2 \times (H-h+1) \times (W-w+1)}$

Parameters:

- Kernel: $\omega \in \mathbb{R}^{C_1 \times C_2 \times h \times w}$
- Bias: (optional) $b \in \mathbb{R}^{C_2}$

$$\underbrace{y_{i,j,k}}_{\text{output}} = \underbrace{b_i}_{\text{bias}} + \sum_{l=1}^{C_1} \sum_{m=0}^{h-1} \sum_{n=0}^{w-1} \underbrace{x_{l,j+m,k+n}}_{\text{input}} \cdot \underbrace{\omega_{i,l,m,n}}_{\text{kernel}}$$



Recap: Convolution

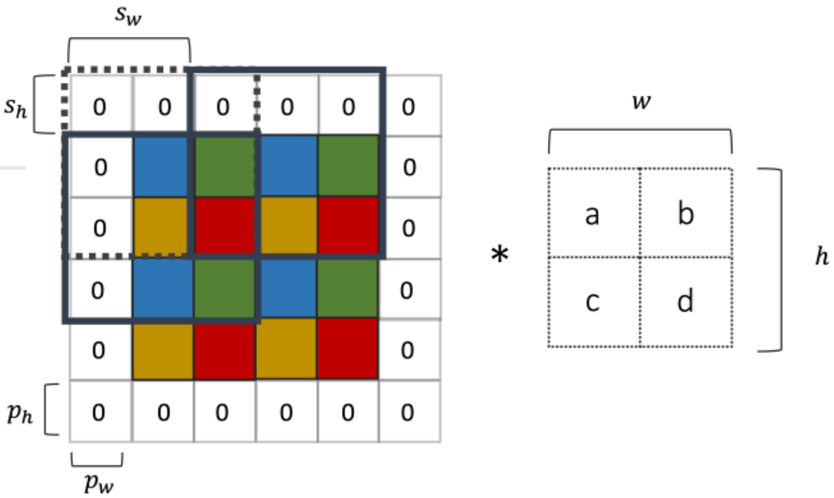
Stride: $s_w = s_h$

Kernel size: $w = h$

Output channels: C_2

Groups: g

Padding: $p_w = \frac{w-1}{2}$ and $p_h = \frac{h-1}{2}$



Convolution - Linear Operator

Convolution is a linear operation *per output* $o_{:,i,j}$

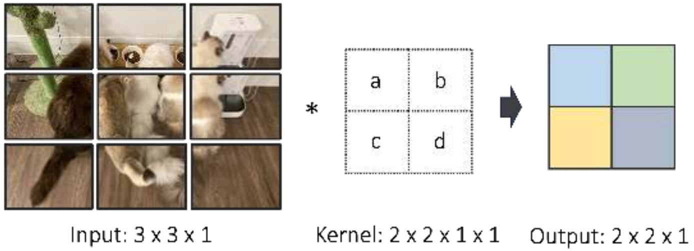
$$o_{:,i,j} = W \mathrm{flatten}(\underbrace{x_{:,i \cdot s_w : i \cdot s_w + w, j \cdot s_h : j \cdot s_h + h}}_{w \times h \text{ image patch}}) + b$$

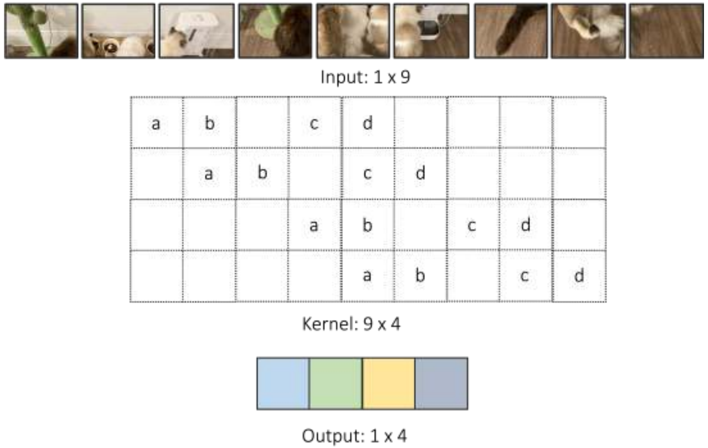
more general

$$o_{:,i,j} = f(\mathrm{flatten}(\underbrace{x_{:,i \cdot s_w : i \cdot s_w + w, j \cdot s_h : j \cdot s_h + h}}_{w \times h \text{ image patch}}))$$

for

$$f(x) = Wx + b$$





Convolutional Operators - Pooling

Slide a function f over image.

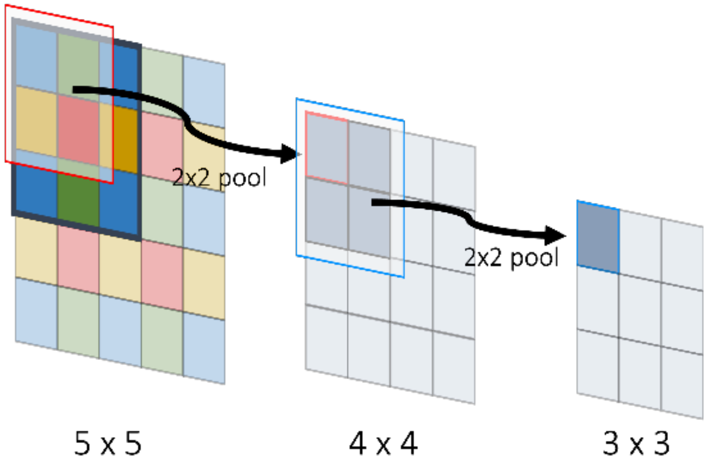
$$o_{:,i,j} = f(\text{flatten}(\underbrace{x_{:,i \cdot s_w : i \cdot s_w + w, j \cdot s_h : j \cdot s_h + h}}_{w \times h \text{ image patch}}))$$

for general f.

Stride: $s_w = s_h$

Kernel size: $w = h$

Padding: $p_w = \frac{w-1}{2}$ and $p_h = \frac{h-1}{2}$

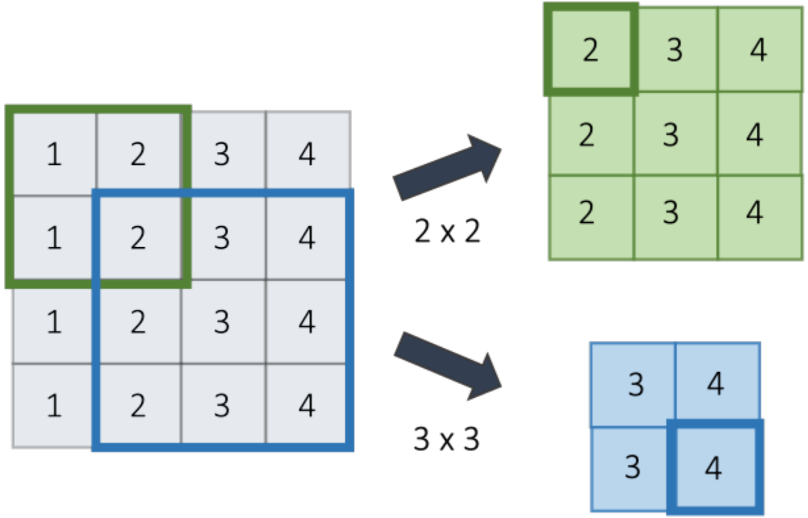


Max Pooling

Slide a channel-wise max $f(x)_c = \max(x_c)$ over image.

$$o_{:,i,j} = f(\text{flatten}(\underbrace{x_{:,i\cdot s_w:i\cdot s_w+w,j\cdot s_h:j\cdot s_h+h}}_{w\times h \text{ image patch}}))$$

- Non-linear



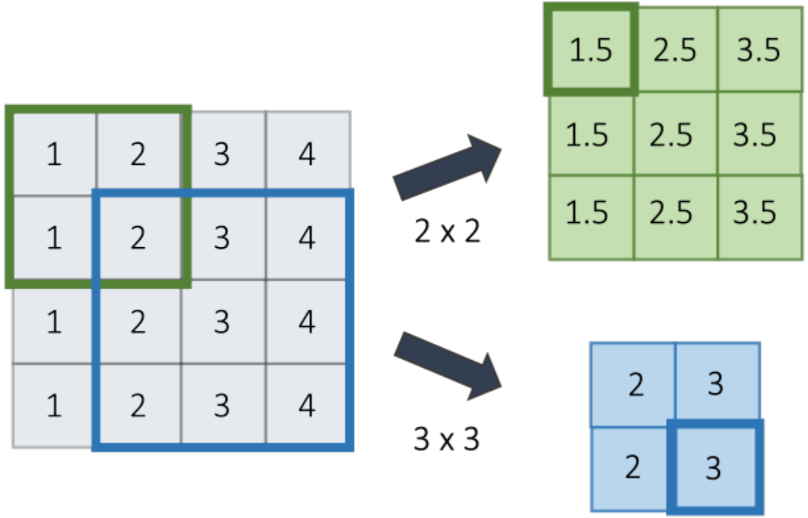
Max pooling

Average Pooling

Slide a channel-wise average function $f(x)_c = \mathrm{avg}(x_c)$ over image.

$$o_{:,i,j} = f(\mathrm{flatten}(\underbrace{x_{:,i\cdot s_w:i\cdot s_w+w,j\cdot s_h:j\cdot s_h+h}}_{w\times h \text{ image patch}}))$$

- Linear operation
- Equivalent to depthwise convolution with
  - $W = \frac{1}{wh}$
  - $b = 0$



Average pooling

Global Pooling

Slide a function f over image.

$$o_{:} = f(\text{flatten}(\underbrace{x}_{\text{image image}}))$$

for general f.

Pooling with kernel size $w = W$ and $h = H$.

No Stride, Kernel size, Padding

Output without spatial dimensions

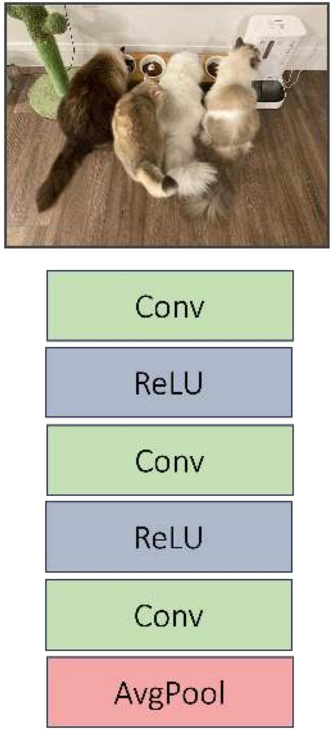
Use Cases

Regular pooling no longer used

- Max Pooling is *historically* used **inside** a network
- Max pool works as a non-linearity layer [1]

Global pooling uses

- Removes spatial dimensions
- Global Average Pooling at the **end**
- Global Max Pooling in PointCloud Processing [2]



Average pooling in the End

Max pooling in the middle

1. Systematic Evaluation of Convolution Neural Network Advances on the ImageNet; Dmytro Mishkin, Nikolay Sergievskiy, Jiri Matas; 2017 ↩
2. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation; Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas; 2016 ↩

Pooling - TL;DR

Convolution-like operator with arbitrary function