Neural Networks: II. Convolutional NN (Part 7)

Jan Bauer

jan.bauer@dhbw-mannheim.de

06.06.19

Motivation

- previous:

$$\underbrace{\text{input}}_{1\times 728} \longrightarrow \underbrace{x \cdot W}_{10\times 728} \longrightarrow \underbrace{\text{score}}_{1\times 10}$$

- now: Keep structure and don't flatten it

$$\rightsquigarrow \text{input is } 28 \times 28 \times 1$$

Convolutional Layer

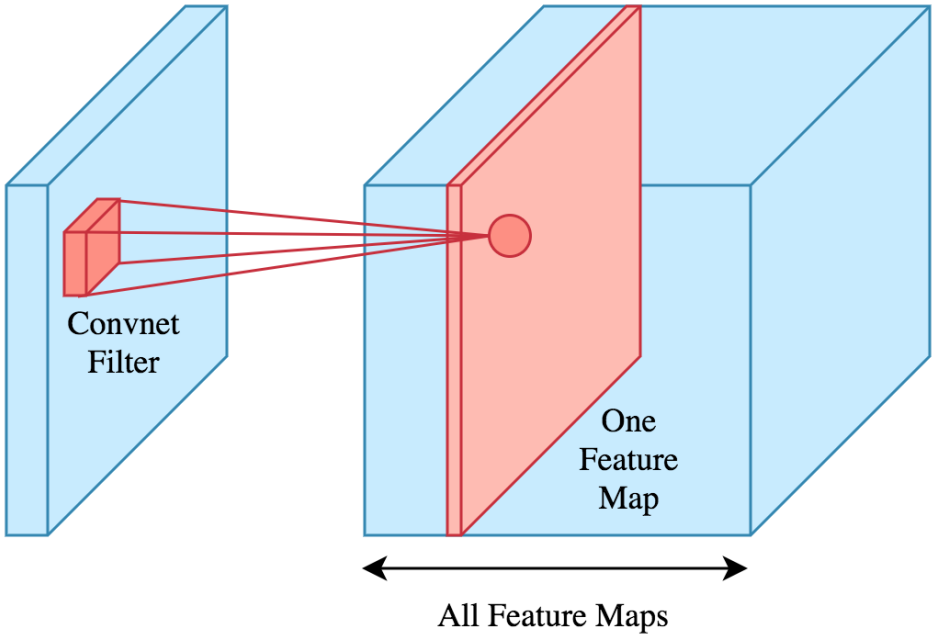

Figure: sliding and projecting (brilliant.org)

- use filter(s) and slide it over the image
- project the slices onto activation map(s) (onto one number):

$$x'w + b$$

- image-depth = filter-depth (foreshadowing)

Convolutional Layer: More elaborate

Example: Input $28 \times 28 \times 1$ and a filter of size $5 \times 5 \times 1$

- use filter(s) and slide it over the image
  - you can use different strides for sliding
- project the slices onto activation map(s) (onto one number):

$$x'w + b \rightsquigarrow 5 \cdot 5 \cdot 1 \text{ dot product}$$

- image-depth = filter-depth
  - example: Coloured images given in RGB i.e. depth = 3

Why a filter?

Intuition:

- the filter has a broader "view" and can scan for particular shapes/features
- the deeper we stand in our network, the more particular are the shapes we are scanning for
- we use several filters, since every filter scans for different shapes
- example: Recognize handwritten digits
  - scanning for a cross
    $\Rightarrow$ 4, 7 or 8
  - scanning for a right angle
    $\Rightarrow$ 4 or 5
  - scanning for a kink
    $\Rightarrow$ 1, 2, 4, 5 or 7
  - scanning for a curve
    $\Rightarrow$ 2, 3, 5, 6, 8, 9 or 0

Example: Dot-product projection when scanning for a feature

Output Size

- Output Size:

$$N - F + 1$$

- Output Size (with stride > 1):

$$\frac{N - F}{\text{stride}} + 1$$

Example: Image 5×5 and filter 3×3

- Running into an issue in here?

Pad the Border

- add a pad containing zeros
- stops dimension shrinking
- pad-size:

$$F = 1 \Rightarrow \text{no zero pads}$$
$$F = 3 \Rightarrow \text{zero pads of } 1$$
$$F = 5 \Rightarrow \text{zero pads of } 2$$

Example: Image 5×5 and filter 3×3 with stride 1

Convolutional NN Architecture

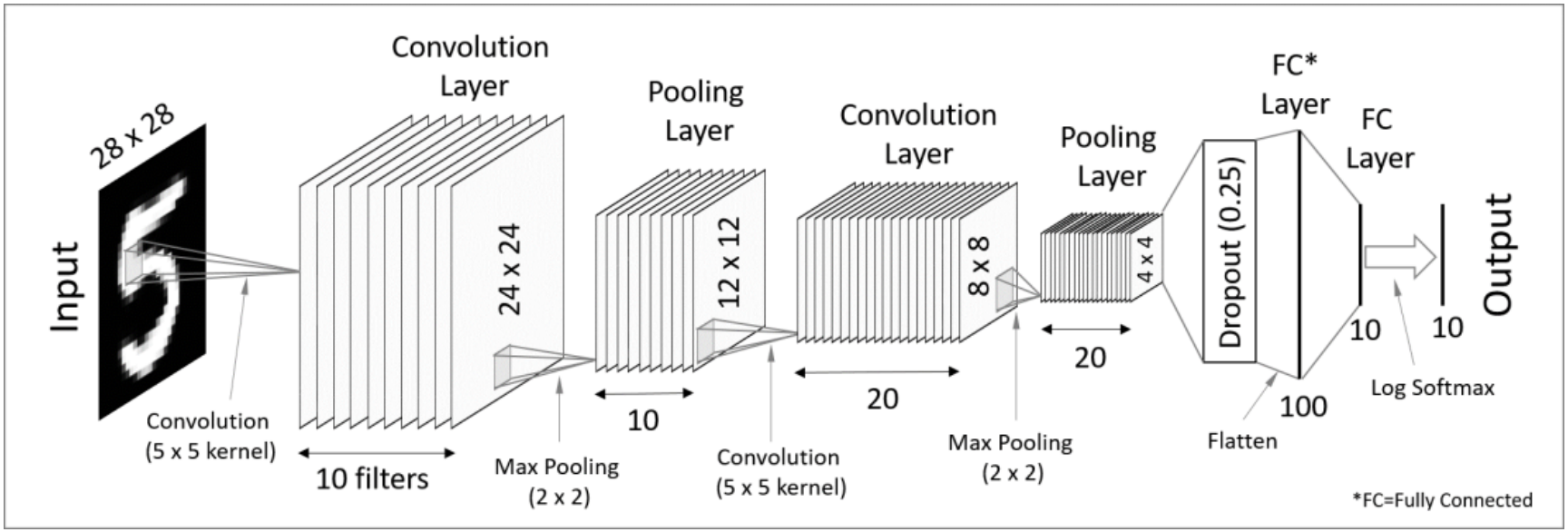

Figure: Architecture (codetolight.files.wordpress.com)

Pooling Layer

- used to reduce the dimensions
- intuition: Focus on the features that have been detected for certain
- apply pooling to every activation map

Example: Max Pooling

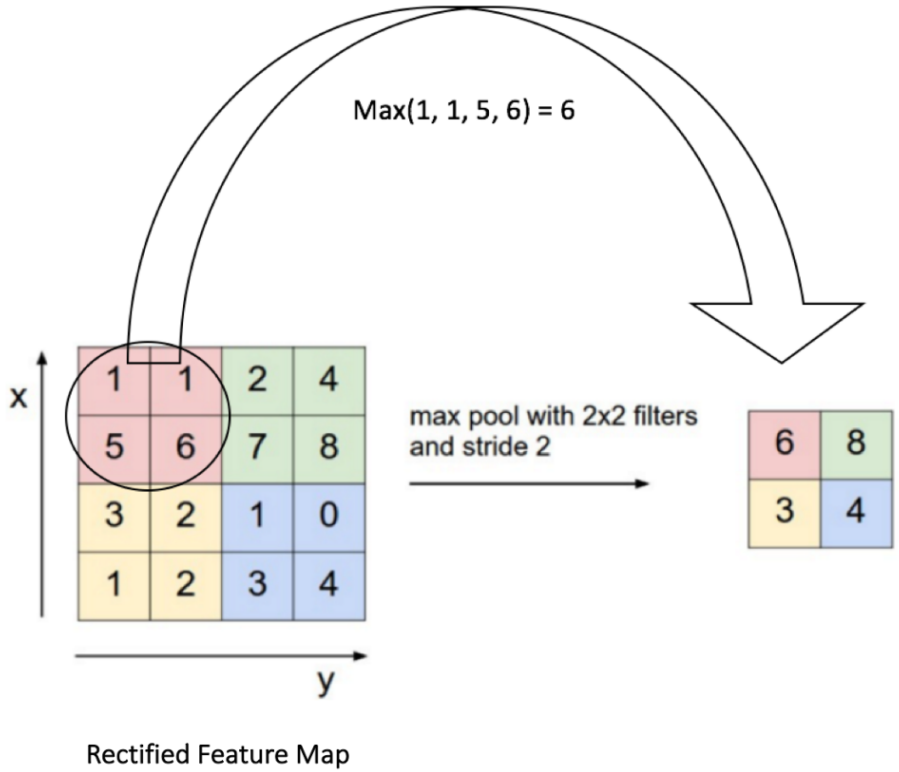

Figure: max pooling (blog.algorithmia.com)