Convolutional Neural Networks

Macroentrenamiento en Inteligencia Artificial (MeIA) 2023

Juan Irving Vásquez
jivg.org

Red de Macrouniversidades de América Latina y el Caribe
Instituto Politécnico Nacional

June 14, 2023©

Convolutional Neural Networks (CNN)

- Position invariance of features

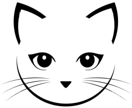
Convolutional Neural Networks (CNN)

- Position invariance of features

- CNNs share their parameters across the space

Template matching

- Template



- Picture

Convolutional Neural Networks insight

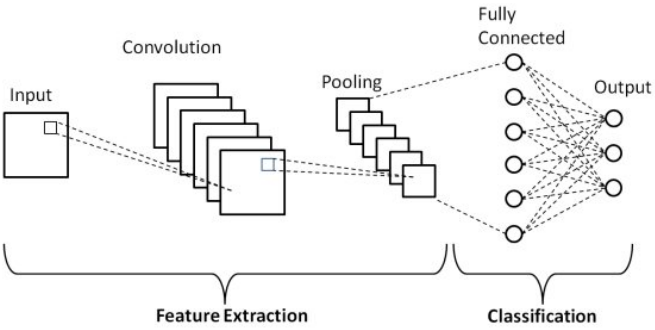


Figure: Scheme of a convolutional neural network for classification.

CNN Input

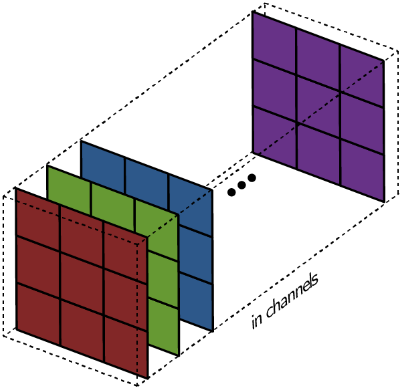


Figure: Illustration of an input image with size $I_w \times I_h \times I_d$.

CNN Convolution

- Given two functions, I and K, the convolution produces a new function that changes the shape of the first one according the to second one.

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} K[u,v] I[i-u, j-v] \tag{1}$$

- It is very useful to find patterns.

CNN Convolution (2)

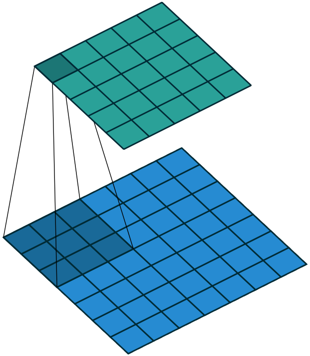
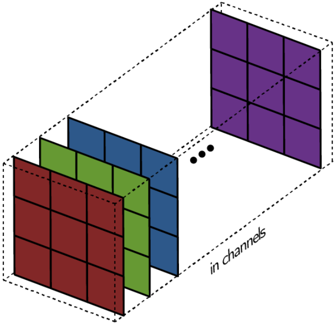


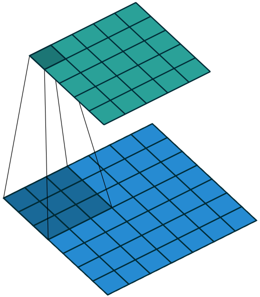
Figure: Convolution: a kernel is moved inside the input to create a new volume.

- The kernel is usually smaller than the input
- The kernel size is $K_w \times K_h \times K_d$

$$K_d = I_d \tag{2}$$



(a) Input image



(b) Convolution

CNN with several Kernels

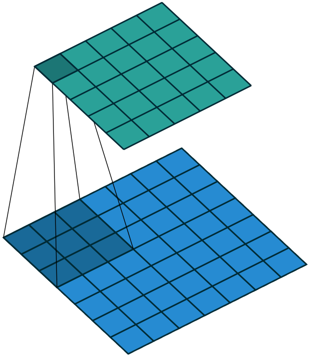
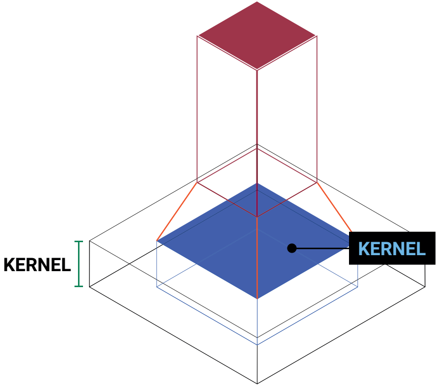


Figure: Several kernels can be applied and their output is stack.

Feature map

- The feature map is the output of the convolution
- The depth of the map, G_d, is equal to the number of filters



CNN Stride

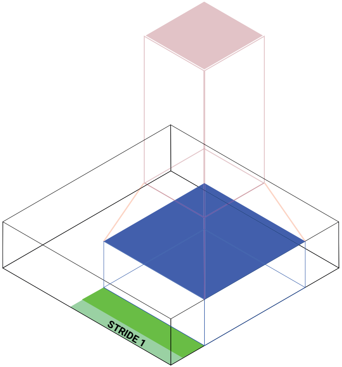


Figure: The stride is the number of pixels that are "jump" for each convolution.

CNN Stride (2)

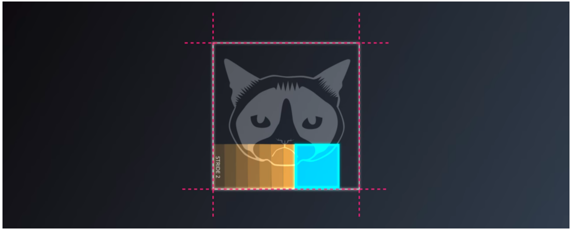
- The stride is defined by s
- The output feature width size is then:

$$G_w = \lfloor \frac{I_w - K_w}{s} \rfloor + 1 \tag{3}$$

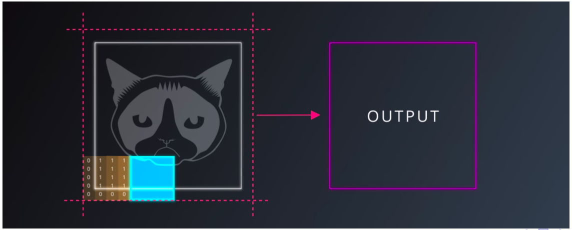
- The same applies for height

CNN Padding

- Valid padding:



- Same padding adds zeros



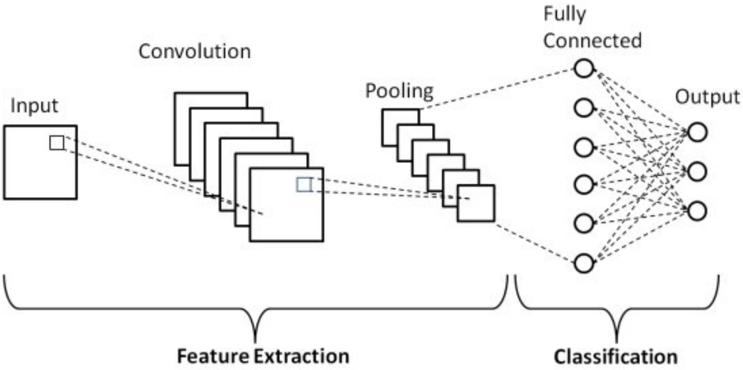


Figure: In a CNN several convolutions are applied before a classification stage.

Exercise

- Given: An input of $28 \times 28 \times 3$, 8 kernels of 3×3, calculate the parameters of the feature maps:

Padding	Stride	Width	Height	Depth
valid	1			
same	1			
valid	2			

Max Pooling

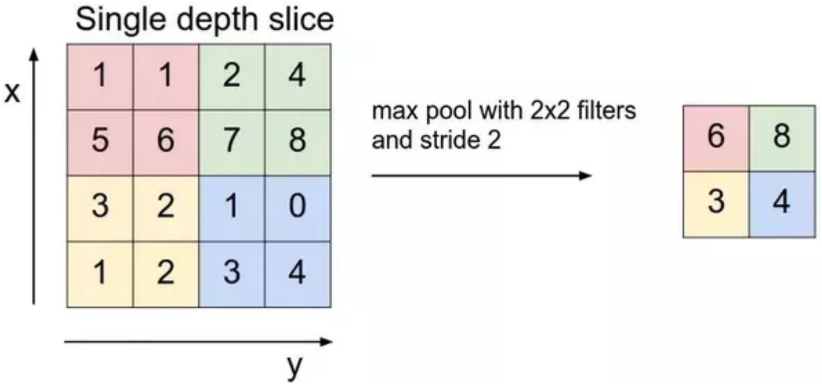


Figure: Pooling reduces the size of the feature map.

Pooling

- Given I_w, W_w and s, the output size is

$$O_w = \lfloor \frac{I_w - W_w}{s} \rfloor + 1 \tag{4}$$

- Some characteristics
  - Weights free
  - A more accurate model
  - Forward pass more expensive
  - More hyper-parameters

1×1 convolution

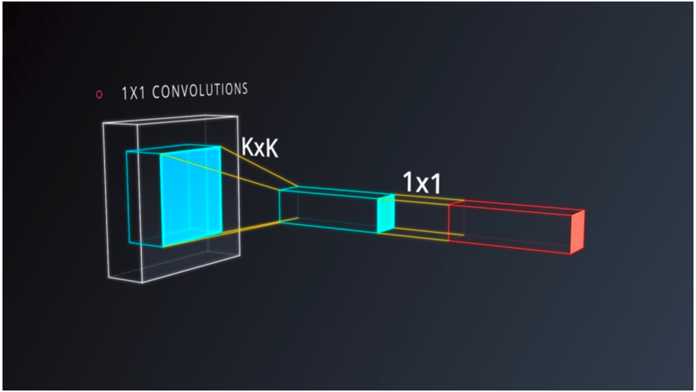


Figure: A 1×1 convolution adds more parameters without modifying the structure.

Inception module (optional)

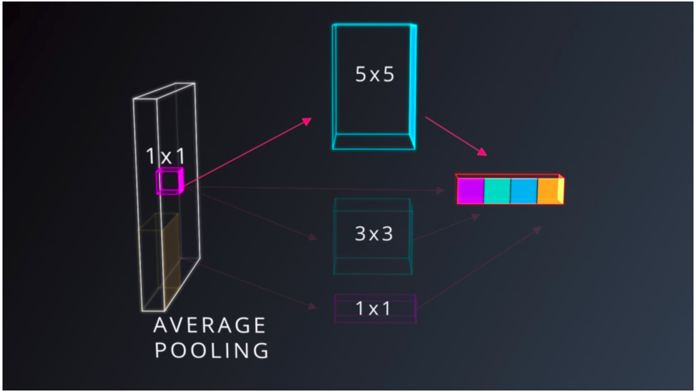


Figure: Why to choose between different operations? Let make all!.

Project - Present

20 minutes presentation

- Lenet5
- AlexNet
- GoogLeNet
- VGGNet

References

- Dumoulin, Vincent, and Francesco Visin. "A guide to convolution arithmetic for deep learning." arXiv preprint arXiv:1603.07285 (2016).
- UDACITY, Computer vision nano degree.
- Skansi, S. (2018). Introduction to Deep Learning: from logical calculus to artificial intelligence. Springer.