Fundamentals of Convolutional Neural Networks Quiz

Anthony Shara
Aditya Ganapathi
DoHyun Cheon
Larry Yan
Richard Shuai

December 11, 2020

1 CNN Advantages

Why would we use convolutional neural networks as opposed to fully connected layers between two feature maps?

2 Weight savings

2.1 Fully connected layer

For an input with dimensions $256 \times 256 \times 3$, calculate the number of weights required for a fully connected layer with 1000 neurons.

2.2 Convolutional layer

If we were to instead use a convolutional layer with filter size 7, sufficient padding and stride to result in an output of size 84x84, and a depth of 32, calculate the number of weights required for this layer.

3 Output size

With input size 32 x 32, a kernel size of 3x3, a stride of 3, and 2 on both sides, what is the result of the output feature map?

4 3x3 Filters

A 3x3 filter covers only 9 neurons while a 15x15 filter covers 225 neurons. How many 3x3 filters are required to achieve the same coverage as a single 15x15 filter? What percentage of weights do we save using only 3x3 filters compared to a 15x15 filter?

5 Same convolution padding

With input size 32 x 32, a kernel size of 7x7, and a stride of 1, what padding is necessary in order to achieve a "same" convolution? (A "same" convolution refers to a convolution which results in an output with the same shape of the original input).

6 Pooling

Why is pooling important in CNN architectures?

7 Convolutions vs Cross correlation

In practice, most CNNs use cross correlation instead of convolutions in their convolutional layers (don't flip filters). Given two implementations of CNNs, one using cross correlation and one using convolutions that are otherwise identical, with weights initialized to zero and given the same training set, what differences would there be in their outputs and filters, if any?

8 Image Classification Intuition

For classification, why do many architectures use fully connected layers after the convolutional layers in order to make classification predictions?

9 Vanishing Gradient Problem

As more layers using ReLU activation functions are added to a CNN, the gradients of a loss function approaches zero. This means that adding more layers to a CNN produces diminishing returns on accuracy, as later layers are unable to learn the function effectively. What technique is used in a famous CNN architecture to combat this vanishing gradient problem?

10 Convolutions as Matrix-Vector Multiplications

Convolutions represent linear transformations, and we have seen how they can be expressed as a matrix vector multiplication for some convolution matrix and some input vector. Let the input image be the 4x4 matrix X given below. We then perform a convolution with a $3x3$ kernel K that is also provided below.

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 0 \\ 3 & 3 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{bmatrix} \qquad K = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

a) Express this convolution as a matrix-vector multiplication $\mathbf{A}\vec{x}$, writing out explicitly what the entries of $\mathbf{A}$ and $\vec{x}$ are. *Hint: Think about what the dimensions of the input and output image will be. What must the dimensions of A be?*

b) Give a reason for why convolutions are not usually implemented as a matrix-vector multiplication in practice.