UNIVERSITÄT DES SAARLANDES

Saarland University, Department of Computer Science
Neural Network Assignment 9

Deborah Dormah Kanubala (7025906) , Irem Begüm Gündüz (7026821), Anh Tuan Tran (7015463)

July 19, 2023

Exercise 9.1 Convolutional Networks

Exercise 9.1.a

(a) Image datasets are represented with pixels. Each pixel corresponds to numeric values. Neural network architectures other than CNN treats pixels by ignoring spatial resolution between them. These networks are invariant to feature order therefore, they may lead to similar results even if we preserve spatial structure between pixels. CNNs convolutes by leveraging features based on a kernel, therefore provides an efficient framework to work with image dataset because it preserves spatial resolution between pixels related to each other.

(b) CNN can be used in other applications, such as classification tasks in natural language processing. Most NLP datasets are large, the convolutional and pooling layers of CNN can help to reduce the size of the dataset by capturing meaningful features. These features then can be used to text classification tasks. However, in other tasks such as part of speech tagging, CNNs outperformed by other algorithms such as bidirectional recurrent neural networks.

Exercise 9.1.b

(a) Write down the equation of how convolution of a given image is computed. Given image I of size $H \times W \times C$, kernel K of size $N \times M \times C \times K$ and stride (T, S), we have output O is

$$O[i, j, k] = \sum_{c=0}^{C} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} I[i * T + n, j * S + m, c] * K[n, m, c, k] \tag{1}$$

(b) Consider a convolutional layer with 10 filters of size (5×5) with, stride =1 and padding = 2. If the input is an image of size $32 \times 32 \times 3$, what will be the output dimensions?

With no information given, we assume that the dilation of the convolutional layer is 1 and the output channel is the same as input channel (3). After one convolution, the width of the output will be:

$$
\begin{aligned}
W_{out} &= \left\lfloor \frac{W_{in} + 2 * \text{padding} - \text{dilation} * (\text{kernel\_size} - 1) - 1}{\text{stride}} + 1 \right\rfloor \\
&= \left\lfloor \frac{32 + 2 * 2 - 1 * (5 - 1) - 1}{1} + 1 \right\rfloor \\
&= 32
\end{aligned} \tag{2}
$$

Similarly, we have $H_{out} = 32$. Therefore, after 10 convolution layer, height and width will be same as input which is (32 x 32). The size of the output will be (32 x 32 x 3)

(c) Calculate the number of parameters in the above layer and compare it to the number of parameters a fully connected layer of the same size.

With no information given, we consider that no bias is used for the convolution layers, output channel of each layer is the same as the input channel (3). The number of learnable parameter for each convolution layer is $kernel\_width * kernel\_height * input\_channel * output\_channel = 5 * 5 * 3 * 3 = 225$

With 10 layers, the total number of parameters is $225 * 10 = 2250$

Considering a single fully connected layer that give the same output size, the input is the flattened image and the output is another flattened image of the same size. The number of parameters is $(image\_width * image\_height * image\_channel)^2 = (32 * 32 * 3)^2 = 9437184$. This is much larger than 10 layers of the said convolution.

Exercise 9.1.c

$$
R_k = R_{k-1} + (K - 1) \prod_{i=1}^{k-1} S_i \tag{3}
$$

(a) Calculate the receptive field at each layer where $K = 5, S = 1$

$$
\begin{aligned}
&R_0 = 1 \\
&R_1 = R_0 + (K - 1) \prod_{i=1}^{k-1} S_i = 1 + 4 * 1 = 5 \\
&R_2 = R_1 + (K - 1) \prod_{i=1}^{k-1} S_i = 5 + 4 * 1 = 9 \\
&\ldots \\
&R_k = R_{k-1} + (K - 1) \prod_{i=1}^{k-1} S_i = 1 + 4 * k
\end{aligned} \tag{4}
$$

(b) How does pooling affect the effective receptive field?

The receptive field size calculation for pooling layer also follows the eq. 3 which dependent on the kernel size, stride and previous receptive field.

References