DS-GA 1008: Homework Assignment 2

Tejaishwarya Gagadam, tg1779

1. Fundamentals

1.1. Convolution

(a) $\boldsymbol{A}$ is a 5×5 single-channel image and $\boldsymbol{B}$ is a 3×3 convolution kernel.
The output (with no padding) $\boldsymbol{C}$ will have the dimensions: $(5-3+1) \times (5-3+1)$. That is, 3×3.

(b) The general formula of the output width is $O \times O$, where $O = \left(\frac{I-K+2P}{S}\right) + 1$. In our example, $I = 5$, $K = 3$, $S = 1$ and $P = 0$.

(c) Assuming that the bias term of the convolution is zero.

$\boldsymbol{C} =$

109	92	72
108	85	74
110	74	79

(d)

$\boldsymbol{A} =$

A_{11}	A_{12}	A_{13}	A_{14}	A_{15}
A_{21}	A_{22}	A_{23}	A_{24}	A_{25}
A_{31}	A_{32}	A_{33}	A_{34}	A_{35}
A_{41}	A_{42}	A_{43}	A_{44}	A_{45}
A_{51}	A_{52}	A_{53}	A_{54}	A_{55}

$\boldsymbol{B} =$

B_{11}	B_{12}	B_{13}
B_{21}	B_{22}	B_{23}
B_{31}	B_{32}	B_{33}

$\boldsymbol{C} =$

C_{11}	C_{12}	C_{13}
C_{21}	C_{22}	C_{23}
C_{31}	C_{32}	C_{33}

The gradient backpropagated from the layers above this layer is a 3×3 matrix of all 1s. Using Chain Rule we have,

$$\frac{\partial E}{\partial A_{i'j'}} = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{\partial E}{\partial C_{ij}} \frac{\partial C_{ij}}{\partial A_{i'j'}}$$

where, $(i', j') \in \{1, 2, 3, 4, 5\}$ and $(i, j) \in \{1, 2, 3\}$

DS-GA 1008: Homework Assignment 2

Tejaishwarya Gagadam, tg1779

- $$\frac{\partial E}{\partial A_{11}} = \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{11}} = B_{11} \tag{1}$$

- $$\frac{\partial E}{\partial A_{12}} = \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{12}} + \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{12}} = B_{11} + B_{12} \tag{2}$$

- $$\frac{\partial E}{\partial A_{13}} = \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{13}} + \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{13}} + \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{13}} = B_{11} + B_{12} + B_{13} \tag{3}$$

- $$\frac{\partial E}{\partial A_{14}} = \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{14}} + \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{14}} = B_{12} + B_{13} \tag{4}$$

- $$\frac{\partial E}{\partial A_{15}} = \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{15}} = B_{13} \tag{5}$$

- $$\frac{\partial E}{\partial A_{21}} = \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{21}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{21}} = B_{11} + B_{21} \tag{6}$$

- $$\begin{multline}\frac{\partial E}{\partial A_{22}} = \\ \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{22}} + \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{22}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{22}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{22}} \\ = B_{11} + B_{12} + B_{21} + B_{22} \tag{7}\end{multline}$$

- $$\begin{multline}\frac{\partial E}{\partial A_{23}} = \\ \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{23}} + \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{23}} + \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{23}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{23}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{23}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{23}} \\ = B_{11} + B_{12} + B_{13} + B_{21} + B_{22} + B_{23} \tag{8}\end{multline}$$

- $$\frac{\partial E}{\partial A_{24}} = \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{24}} + \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{24}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{24}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{24}} = B_{12} + B_{13} + B_{22} + B_{23} \tag{9}$$

- $$\frac{\partial E}{\partial A_{25}} = \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{25}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{25}} = B_{13} + B_{23} \tag{10}$$

- $$\frac{\partial E}{\partial A_{31}} = \frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{31}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{31}} + \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{31}} = B_{11} + B_{21} + B_{31} \tag{11}$$

- $$\begin{aligned}\frac{\partial E}{\partial A_{32}} &= \\ &\frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{32}} + \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{32}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{32}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{32}} + \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{32}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{32}} \\ &= B_{11} + B_{12} + B_{21} + B_{22} + B_{31} + B_{32}\end{aligned} \tag{12}$$

- $$\begin{aligned}\frac{\partial E}{\partial A_{33}} &= \\ &\frac{\partial E}{\partial C_{11}}\frac{\partial C_{11}}{\partial A_{33}} + \frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{33}} + \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{33}} + \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{32}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{33}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{33}} \\ &+ \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{33}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{33}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{33}} \\ &= B_{11} + B_{12} + B_{13} + B_{21} + B_{22} + B_{23} + B_{31} + B_{32} + B_{33}\end{aligned} \tag{13}$$

- $$\begin{aligned}\frac{\partial E}{\partial A_{34}} &= \\ &\frac{\partial E}{\partial C_{12}}\frac{\partial C_{12}}{\partial A_{34}} + \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{34}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{34}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{34}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{34}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{34}} \\ &= B_{12} + B_{13} + B_{22} + B_{23} + B_{32} + B_{33}\end{aligned} \tag{14}$$

- $$\frac{\partial E}{\partial A_{35}} = \frac{\partial E}{\partial C_{13}}\frac{\partial C_{13}}{\partial A_{35}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{35}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{35}} = B_{13} + B_{23} + B_{33} \tag{15}$$

- $$\frac{\partial E}{\partial A_{41}} = \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{41}} + \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{41}} = B_{21} + B_{31} \tag{16}$$

- $$\frac{\partial E}{\partial A_{42}} = \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{42}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{42}} + \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{42}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{42}} = B_{21} + B_{22} + B_{31} + B_{32} \tag{17}$$

DS-GA 1008: Homework Assignment 2

Tejaishwarya Gagadam, tg1779

- $$\frac{\partial E}{\partial A_{43}} = \frac{\partial E}{\partial C_{21}}\frac{\partial C_{21}}{\partial A_{42}} + \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{42}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{42}} + \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{42}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{42}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{42}} = B_{21} + B_{22} + B_{23} + B_{31} + B_{32} + B_{33} \tag{18}$$

- $$\frac{\partial E}{\partial A_{44}} = \frac{\partial E}{\partial C_{22}}\frac{\partial C_{22}}{\partial A_{44}} + \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{44}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{44}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{44}} = B_{22} + B_{23} + B_{32} + B_{33} \tag{19}$$

- $$\frac{\partial E}{\partial A_{45}} = \frac{\partial E}{\partial C_{23}}\frac{\partial C_{23}}{\partial A_{45}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{45}} = B_{23} + B_{33} \tag{20}$$

- $$\frac{\partial E}{\partial A_{51}} = \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{51}} = B_{31} \tag{21}$$

- $$\frac{\partial E}{\partial A_{52}} = \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{52}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{52}} = B_{31} + B_{32} \tag{22}$$

- $$\frac{\partial E}{\partial A_{53}} = \frac{\partial E}{\partial C_{31}}\frac{\partial C_{31}}{\partial A_{53}} + \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{53}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{53}} = B_{31} + B_{32} + B_{33} \tag{23}$$

- $$\frac{\partial E}{\partial A_{54}} = \frac{\partial E}{\partial C_{32}}\frac{\partial C_{32}}{\partial A_{54}} + \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{54}} = B_{32} + B_{33} \tag{24}$$

- $$\frac{\partial E}{\partial A_{55}} = \frac{\partial E}{\partial C_{33}}\frac{\partial C_{33}}{\partial A_{55}} = B_{33} \tag{25}$$

For all the equations above, we have

$\frac{\partial E}{\partial A_{11}}$	$\frac{\partial E}{\partial A_{12}}$	$\frac{\partial E}{\partial A_{13}}$	$\frac{\partial E}{\partial A_{14}}$	$\frac{\partial E}{\partial A_{15}}$
$\frac{\partial E}{\partial A_{21}}$	$\frac{\partial E}{\partial A_{22}}$	$\frac{\partial E}{\partial A_{23}}$	$\frac{\partial E}{\partial A_{24}}$	$\frac{\partial E}{\partial A_{25}}$
$\frac{\partial E}{\partial A_{31}}$	$\frac{\partial E}{\partial A_{32}}$	$\frac{\partial E}{\partial A_{33}}$	$\frac{\partial E}{\partial A_{34}}$	$\frac{\partial E}{\partial A_{35}}$
$\frac{\partial E}{\partial A_{41}}$	$\frac{\partial E}{\partial A_{42}}$	$\frac{\partial E}{\partial A_{43}}$	$\frac{\partial E}{\partial A_{44}}$	$\frac{\partial E}{\partial A_{45}}$
$\frac{\partial E}{\partial A_{51}}$	$\frac{\partial E}{\partial A_{52}}$	$\frac{\partial E}{\partial A_{53}}$	$\frac{\partial E}{\partial A_{54}}$	$\frac{\partial E}{\partial A_{55}}$

=

4	7	10	6	3
9	17	25	16	8
11	23	35	23	11
7	16	24	17	8
2	6	9	7	3

DS-GA 1008: Homework Assignment 2

Tejaishwarya Gagadam, tg1779

1.2. Pooling

Pooling is a technique for sub-sampling and comes in different flavors, for example max-pooling, average pooling, LP-pooling.

(a) The `torch.nn` modules for the 2D versions of these pooling techniques are:

- Max-Pooling: `torch.nn.module.MaxPool2d`
- Average Pooling: `torch.nn.module.AvgPool2d`
- LP-pooling: `torch.nn.module.LPPool2d`

(b) The k-th input feature maps to a pooling module as $\boldsymbol{X}^k \in \mathbb{R}^{H_{\text{in}} \times W_{\text{in}}}$ where H_{in} and W_{in} represent the input height and width, respectively. Let $\boldsymbol{Y}^k \in \mathbb{R}^{H_{\text{out}} \times W_{\text{out}}}$ denote the k-th output feature map of the module where H_{out} and W_{out} represent the output height and width, respectively. Let $S^k_{i,j}$ be a list of the indexes of elements in the sub-region of X^k used for generating $\mathbf{Y}^k_{i,j}$, the (i, j)-th entry of $\boldsymbol{Y}^k$. Then we have,

- **Max-Pooling:**
  $\mathbf{Y}^k_{i,j} = max\{\boldsymbol{X}^k_{i',j'}\}$, where $(i', j') \in S^k_{i,j}$
- **Average Pooling:**
  $\mathbf{Y}^k_{i,j} = \frac{\sum_{i'}\sum_{j'}\{\boldsymbol{X}^k_{i',j'}\}}{N}$, where $(i', j') \in S^k_{i,j}$, and N = Number of elements in the sub-region of X^k
- **LP-Pooling:**
  $\mathbf{Y}^k_{i,j} = \left(\frac{\sum_{i'}\sum_{j'}\{\boldsymbol{X}^k_{i',j'}\}^P}{N}\right)^{1/P}$, where $(i', j') \in S^k_{i,j}$, and N = Number of elements in the sub-region of X^k
- Note: If $S^k_{i,j}$ could be represented as a matrix of $z \times z$, then $z^2 = N$

(c) Applying max-pooling module with kernel size of 2 and stride of 1 to $\boldsymbol{C}$, we get

109	92
110	85

(d) Max-pooling and average pooling can be expressed in terms of LP-pooling, in the following way

$$\boldsymbol{Y}^k_{i,j} = \left(\frac{\sum_{i'}\sum_{j'}\{\boldsymbol{X}^k_{i',j'}\}^P}{N}\right)^{1/P} \begin{cases} \text{when p} = 1 \text{ Average Pooling,} \\ \text{when p} \rightarrow \infty \text{ Max-Pooling} \end{cases}$$