A Appendix: Convolution Backwards Derivation

Here we will derive the convolution gradient formulars used in Question 1. To prove this, we will derive the gradients of the filter convolution then use the chain rule to find the final convolution gradient. First, we will first introduce some indexing notation, this will help cut down on the writing and make it easier to see what's going on.

Notation. Let $\mathbf{m} \in \mathbb{R}^{H_m \times W_m}$ be some matrix. We define the "slice" of $\mathbf{m}$ as

$$(\mathbf{m})_{i,j}^{k,\ell}$$

which denotes a matrix of size $(k-i) \times (\ell - j)$, where its elements are defined as

$$\left[(\mathbf{m})_{i,j}^{k,\ell}\right]_{n,m} = \begin{cases} \mathbf{m}_{i+n-1,j+m-1} & \text{if } i+n-1, j+m-1 \text{ is contained in } \mathbf{m} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Intuitively, the slice of $\mathbf{m}$, $(\mathbf{m})_{i,j}^{k,\ell}$ can be thought of as zero-padding $\mathbf{m}$, then taking the i, j to k, ℓ indexes of the padded $\mathbf{m}$.

Consider the 2D arrays/tensors defined the same way as Question: $\mathbf{a} \in \mathbb{R}^{H_a \times W_a}$ and $\mathbf{b} \in \mathbb{R}^{H_b \times W_b}$. Let f be any scalar output function that takes $\mathbf{a} *_{\text{filt}} \mathbf{b}$ as input.

With this notation, we can show the following:

Claim:

$$\frac{\partial(\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{a}} = (\mathbf{b})_{2-i,2-j}^{H_a-i+1,W_a-j+1} \tag{2}$$

Proof.

$$\frac{\partial(\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial a_{k,\ell}} = \frac{\partial}{\partial a_{k,\ell}} \sum_{m=i}^{i+H_b-1} \sum_{n=j}^{j+W_b-1} a_{m,n} b_{m-i+1,n-j+1} \tag{3}$$

$$= \sum_{m=i}^{i+H_b-1} \sum_{n=j}^{j+W_b-1} \frac{\partial}{\partial a_{k,\ell}} a_{m,n} b_{m-i+1,n-j+1} \tag{4}$$

$$= \sum_{m=i}^{i+H_b-1} \sum_{n=j}^{j+W_b-1} b_{m-i+1,n-j+1} \mathbb{1}[m = k, n = \ell] \tag{5}$$

$$= b_{k-i+1,\ell-j+1} \tag{6}$$

$$\frac{\partial(\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{a}} = [b_{k-i+1,\ell-j+1}]_{k=1\ldots H_a,\ell=1\ldots W_a} \tag{7}$$

$$= (\mathbf{b})_{1-i+1,1-j+1}^{H_a-i+1,W_a-j+1} \tag{8}$$

$$= (\mathbf{b})_{2-i,2-j}^{H_a-i+1,W_a-j+1} \tag{9}$$

Next, in a similar manner:

Claim:

$$\frac{\partial(\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{b}} = (\mathbf{a})_{i,j}^{i+H_b-1, j+W_b-1} \tag{10}$$

Proof.

$$\begin{aligned}
\frac{\partial(\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial b_{k,\ell}} &= \frac{\partial}{\partial b_{k,\ell}} \sum_{p=1}^{H_b} \sum_{q=1}^{W_b} a_{i+p-1,j+q-1} b_{p,q} && (11)\\
&= \sum_{p=1}^{H_b} \sum_{q=1}^{W_b} \frac{\partial}{\partial b_{k,\ell}} a_{i+p-1,j+q-1} b_{p,q} && (12)\\
&= \sum_{p=1}^{H_b} \sum_{q=1}^{W_b} a_{i+p-1,j+q-1} \mathbb{1}[p=k, q=\ell] && (13)\\
&= a_{i+k-1,j+\ell-1} && (14)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial(\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{b}} &= \left[a_{i+k-1,j+\ell-1}\right]_{k=1\ldots H_b, \ell=1\ldots W_b} && (15)\\
&= (\mathbf{a})_{i+1-1,j+1-1}^{i+H_b-1,j+W_b-1} && (16)\\
&= (\mathbf{a})_{i,j}^{i+H_b-1,j+W_b-1} && (17)
\end{aligned}$$

Next, we can show the following (which can be directly used with the chain rule to find the convolution gradients).

Claim:

$$\frac{\partial f}{\partial \mathbf{a}} = \mathbf{b} *_{\text{full}} \frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}} \tag{18}$$

Proof. Let $H' = H_a - H_b + 1$ and $W' = W_a - W_b + 1$ (the size of the output of $\mathbf{a} *_{\text{filt}} \mathbf{b}$).

$$\begin{aligned}
\left[\frac{\partial f}{\partial \mathbf{a}}\right]_{k,\ell} &= \left[\sum_{i=1}^{H'}\sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} \frac{\partial (\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{a}}\right]_{k,\ell} && (19)\\
&= \sum_{i=1}^{H'}\sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} \left(\frac{\partial (\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{a}}\right)_{k,\ell} && (20)\\
&= \sum_{i=1}^{H'}\sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} \left((\mathbf{b})_{2-i,2-j}^{H_a-i+1,W_a-j+1}\right)_{k,\ell} && (21)\\
&= \sum_{i=1}^{H'}\sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} (\mathbf{b})_{2-i+k-1,2-j+\ell-1} && (22)\\
&= \sum_{i=1}^{H'}\sum_{j=1}^{W'} (\mathbf{b})_{k-i+1,\ell-j+1} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} && (23)\\
&= \sum_{m=k-H'+1}^{k}\ \sum_{n=\ell-W'+1}^{\ell} (\mathbf{b})_{m,n} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{k-m+1,\ell-n+1} && (24)\\
&= \left(\mathbf{b} *_{\text{full}} \frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{k,\ell} && (25)
\end{aligned}$$

and the following:

Claim:

$$\frac{\partial f}{\partial \mathbf{b}} = \mathbf{a} *_{\text{filt}} \frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}} \tag{26}$$

Proof. Let $H' = H_a - H_b + 1$ and $W' = W_a - W_b + 1$ (the size of the output of $\mathbf{a} *_{\text{filt}} \mathbf{b}$).

$$\left[\frac{\partial f}{\partial \mathbf{b}}\right]_{k,\ell} = \left[\sum_{i=1}^{H'} \sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} \frac{\partial (\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{b}}\right]_{k,\ell} \tag{27}$$

$$= \sum_{i=1}^{H'} \sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} \left(\frac{\partial (\mathbf{a} *_{\text{filt}} \mathbf{b})_{i,j}}{\partial \mathbf{b}}\right)_{k,\ell} \tag{28}$$

$$= \sum_{i=1}^{H'} \sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} \left((\mathbf{a})_{i,j}^{i+H'-1,j+W'-1}\right)_{k,\ell} \tag{29}$$

$$= \sum_{i=1}^{H'} \sum_{j=1}^{W'} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} (\mathbf{a})_{i+k-1,j+\ell-1} \tag{30}$$

$$= \sum_{i=1}^{H'} \sum_{j=1}^{W'} (\mathbf{a})_{i+k-1,j+\ell-1} \left(\frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{i,j} \tag{31}$$

$$= \left(\mathbf{a} *_{\text{filt}} \frac{\partial f}{\partial \mathbf{a} *_{\text{filt}} \mathbf{b}}\right)_{k,\ell} \tag{32}$$

Given Equations 18 and 26, we can simply derive the convolution gradient using the chain rule. Recall that

$$Y_{n,f} = \sum_{c}^{C} X_{n,c} *_{\text{valid}} \overline{K}_{f,c} = \sum_{c}^{C} X_{n,c} *_{\text{filt}} K_{f,c} \tag{33}$$

As L is a scalar valued function which takes $Y_{n,f}$ as input,

$$\frac{\partial L}{\partial X_{n,c}} = \sum_{f=1}^{F} \frac{\partial L}{\partial Y_{n,f}} \frac{\partial Y_{n,f}}{\partial X_{n,c}} = \sum_{f=1}^{F} K_{f,c} *_{\text{full}} \frac{\partial L}{\partial Y_{n,f}} \tag{34}$$

and

$$\frac{\partial L}{\partial K_{f,c}} = \sum_{n=1}^{N} \frac{\partial L}{\partial Y_{n,f}} \frac{\partial Y_{n,f}}{\partial K_{f,c}} = \sum_{n=1}^{N} X_{n,c} *_{\text{filt}} \frac{\partial L}{\partial Y_{n,f}}. \tag{35}$$