Neural Networks' Convolution in terms of Tensor products and contractions

Loizos Shianios

February 4, 2019

We will write the operation of a convolutional layer in terms of tensors, in aim to make it easier to compute the gradient of the layer symbolically. We will derive the result through a worked example. Consider a 5×5 matrix to be convolved with a 3×3 filter,

$$\begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{bmatrix} * \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}.$$

Now consider the R_{11} element of the resulting matrix.

$R_{11} = a \times f_{11} + b \times f_{12} + c \times f_{13} + f \times f_{21} + g \times f_{22} + h \times f_{23} + k \times f_{31} + l \times f_{32} + m \times f_{33}$.

Essentially what we are doing is picking a 3×3 matrix from the data matrix and multiply element-wise with the filter matrix and get the sum of the products. We need to break this operation to a series of tensor products and contractions. But first, how can we pick that upper left matrix. Well, we can multiply by a 3×5 matrix on the left and a 5×3 matrix on the right.

$$\begin{bmatrix} ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \end{bmatrix} \cdot \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{bmatrix} \cdot \begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix} = \begin{bmatrix} a & b & c \\ f & g & h \\ k & l & m \end{bmatrix}.$$

The obvious numbers to use are 1 and 0, in an identity like structure.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{bmatrix} = \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \end{bmatrix},$$

$$\begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} a & b & c \\ f & g & h \\ k & l & m \end{bmatrix}.$$

How about, this part,

$$\begin{bmatrix} b & c & d \\ g & h & i \\ l & m & n \end{bmatrix}?$$

We need to move the identity matrix by one step down in the right matrix.

$$\begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} b & c & d \\ g & h & i \\ l & m & n \end{bmatrix}.$$

And how about

$$\begin{bmatrix} m & n & o \\ r & s & t \\ w & x & y \end{bmatrix}?$$

Well, we need to move the identity parts in both matrices to the 'end'.

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{bmatrix} = \begin{bmatrix} k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{bmatrix},$$

$$\begin{bmatrix} k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} m & n & o \\ r & s & t \\ w & x & y \end{bmatrix}.$$

Observe that there are three possible structures for both matrices. Let's call the left matrix X and the right matrix Y. So for X we have

$$X_{1ij} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, X_{2ij} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, X_{3ij} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

And for Y we have,

$$Y_{1kl} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, Y_{2kl} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, Y_{3kl} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that, matrix multiplication is basically the tensor product of two rank 2 tensors followed by the contraction of the two middle indices; that is

$$\boldsymbol{A}_{ij} \cdot \boldsymbol{B}_{jk} = \sum_{j}^{|j|} \boldsymbol{A}_{ij} \cdot \boldsymbol{B}_{jk} = \boldsymbol{A}_{i1}\boldsymbol{B}_{1k} + \boldsymbol{A}_{i2}\boldsymbol{B}_{2k} + \cdots = C_{ik}$$

Let's expand the above equation and see what those $\boldsymbol{A}_{ij}\boldsymbol{B}_{jk}$ terms mean. Consider the first matrix multiplication we had above with the $\boldsymbol{X}_{1ij}$ and the 5×5 matrix, let's call it $\boldsymbol{D}_{jk}$. Then, $|j| = 5$ and we suppress the 1 index of $\boldsymbol{X}$, for now,

$$\sum_{j}^{|j|} \boldsymbol{X}_{ij} \cdot \boldsymbol{D}_{jk} = \boldsymbol{X}_{i1}\boldsymbol{D}_{1k} + \boldsymbol{X}_{i2}\boldsymbol{D}_{2k} + \boldsymbol{X}_{i3}\boldsymbol{D}_{3k} + \boldsymbol{X}_{i4}\boldsymbol{D}_{4k} + \boldsymbol{X}_{i4}\boldsymbol{D}_{4k} + \boldsymbol{X}_{i5}\boldsymbol{D}_{5k}$$

$$\boldsymbol{X}_{i1}\boldsymbol{D}_{1k} = \begin{bmatrix} X_{11}D_{11} & X_{11}D_{12} & X_{11}D_{13} & X_{11}D_{14} & X_{11}D_{15} \\ X_{21}D_{11} & X_{21}D_{12} & X_{21}D_{13} & X_{21}D_{14} & X_{21}D_{15} \\ X_{31}D_{31} & X_{31}D_{12} & X_{31}D_{13} & X_{31}D_{14} & X_{31}D_{15} \end{bmatrix} = \begin{bmatrix} 1a & 1b & 1c & 1d & 1e \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{X}_{i2}\boldsymbol{D}_{2k} = \begin{bmatrix} X_{12}D_{21} & X_{12}D_{22} & X_{12}D_{23} & X_{12}D_{24} & X_{12}D_{25} \\ X_{22}D_{21} & X_{22}D_{22} & X_{22}D_{23} & X_{22}D_{24} & X_{22}D_{25} \\ X_{32}D_{31} & X_{32}D_{22} & X_{32}D_{23} & X_{32}D_{24} & X_{32}D_{25} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ f & g & h & i & j \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{X}_{i3}\boldsymbol{D}_{3k} = \begin{bmatrix} X_{13}D_{31} & X_{13}D_{32} & X_{13}D_{33} & X_{13}D_{34} & X_{13}D_{35} \\ X_{23}D_{31} & X_{23}D_{32} & X_{23}D_{33} & X_{23}D_{34} & X_{23}D_{35} \\ X_{33}D_{31} & X_{33}D_{32} & X_{33}D_{33} & X_{33}D_{34} & X_{33}D_{35} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ k & l & m & n & o \end{bmatrix}$$

$$\boldsymbol{X}_{i4}\boldsymbol{D}_{4k} = \begin{bmatrix} X_{14}D_{41} & X_{14}D_{42} & X_{14}D_{43} & X_{14}D_{44} & X_{14}D_{45} \\ X_{24}D_{41} & X_{24}D_{42} & X_{24}D_{43} & X_{24}D_{44} & X_{24}D_{45} \\ X_{24}D_{41} & X_{24}D_{42} & X_{24}D_{43} & X_{24}D_{44} & X_{24}D_{45} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{X}_{i5}\boldsymbol{D}_{5k} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Adding all the matrices together we get,

$$\boldsymbol{A}_{ij}\boldsymbol{B}_{jk} = \begin{bmatrix} a & b & c & d & e \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ f & g & h & i & j \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ k & l & m & n & o \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \end{bmatrix}$$

So we can write the selection of the matrix part as $\boldsymbol{X}_{aij}\boldsymbol{D}_{jk}\boldsymbol{Y}_{bkl} = \boldsymbol{A}_{aibl}$. And in a similar fashion we can write the operation of the filter matrix with this result, $\boldsymbol{A}_{aibl}\boldsymbol{F}_{il} = \boldsymbol{R}_{ab}$. So the full operation is given by $\boldsymbol{X}_{aij}\boldsymbol{D}_{jk}\boldsymbol{Y}_{bkl}\boldsymbol{F}_{il} = \boldsymbol{R}_{ab}$.

Note that if stride is used then the structure of the $\boldsymbol{X}$ and $\boldsymbol{Y}$ will change. The identity parts will move stride steps instead of one step. For example if stride is set to 2, in the previous example then,

$$X_{1ij} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \qquad X_{2ij} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$Y_{1kl} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad Y_{2kl} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

So our expression remains the same but now the indices a, b run from 1 to 2 instead from 1 to 3. Also observe that the structure of the i, l indices are determine by the filter structure whereas, the structure of the j, k indices is determine by the data matrix.

Finally we will need to consider padding and convolutions over volume or higher rank tensor. More on this in another update.