Lecture 6 - The Convolutional Layer in Convolutional Networks

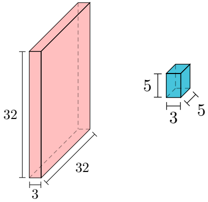
DD2424

April 19, 2017

ConvNets for RGB Images: The Convolution Layer

Convolution Layer

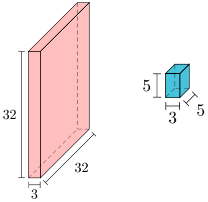
Input Image **Filter**



X is $32 \times 32 \times 3$ F is $5 \times 5 \times 3$

Convolution Layer

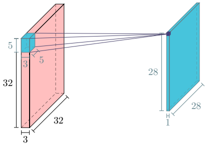
Input Image **Filter**



X is $32 \times 32 \times 3$ F is $5 \times 5 \times 3$

Note: Filter & input image always have the same depth.

Convolution Layer



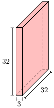
Convolve the image, X, with the filter F.

- Slide filter over all spatial locations in image.
- At each location output 1 number:

  compute **dot product** *between* F *and a* $5 \times 5 \times 3$ *chunk of* X

Convolution Layer

Input Image



X

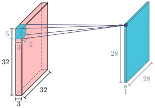
Size: $32 \times 32 \times 3$

Filter

F

Size: $5 \times 5 \times 3$

Output response map



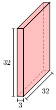
$S = X * F$

Size: $28 \times 28 \times 1$

Convolution Layer

Can apply multiple filters.

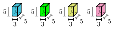
Input Image



X

Size: $32 \times 32 \times 3$

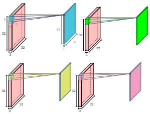
Filters



F_1, F_2, F_3, F_4

Size F_i: $5 \times 5 \times 3$

Output response maps



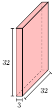
$S_i = X * F_i$

Size S_i: $28 \times 28 \times 1$

Convolution Layer

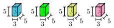
Apply multiple filters and get multiple response maps

Input Image



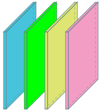
X

Filters



F_1, F_2, F_3, F_4

Output response maps

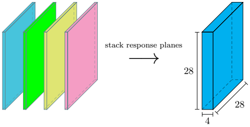


S_1, S_2, S_3, S_4

Each $S_i = X * F_i$

Convolution Layer

- Stack the multiple response maps to get a *new image* S.
- In our example
  - $S = \{S_1, S_2, S_3, S_4\}$ and
  - S has size $28 \times 28 \times 4$



Convolution Layer

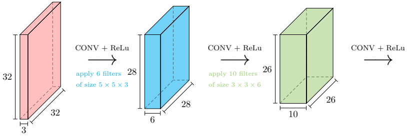
- Apply the non-linear activation function to each element of S.

$$H = \max(0, S)$$

Most basic Convolutional Network layers

Basic **ConvNet** is a composition of

- Convolution Layer
- Activation function



How do we produce final probs for C class labels?

- Add **fully connected layer(s)** after the convolutional layers.
- Example network:
  1 convolutional layer + 1 fully connected layer

$$S_i = X * F_i + b_i \quad \text{for } i = 1, \ldots, n_F \quad \leftarrow \text{apply convolution filters}$$

$$S = \{S_1, \ldots, S_{n_F}\} \quad \leftarrow \text{stack response maps, get new 3D image}$$

$$H = \max(0, S) \quad \leftarrow \text{apply ReLu}$$

$$\mathbf{s} = W\text{vec}(H) + \mathbf{b} \quad \leftarrow \text{fully-connected layer to get } C \text{ scores}$$

$$\mathbf{p} = \text{SOFTMAX}(\mathbf{s}) \quad \leftarrow \text{turn scores into probabilities}$$

- Dimensions of inputs, outputs and parameters:
  - X is $w \times h \times 3$
  - Each F_i is $f \times f \times 3$ and b_i is a scalar
  - Each S_i is $(w - f + 1) \times (h - f + 1)$
  - S and H are $(w - f + 1) \times (h - f + 1) \times n_F$
  - W is $C \times (w - f + 1)(h - f + 1)n_F$
  - $\mathbf{b}, \mathbf{s}$ and $\mathbf{p}$ are $C \times 1$

How do we learn the parameters of the network?

- Add **fully connected layer(s)** after the convolutional layers.
- Example network:
  1 convolutional layer + 1 fully connected layer

  $S_i = X * F_i + b_i$ for $i = 1, \ldots, n_F$ $\leftarrow$ apply convolution filters

  $S = \{S_1, \ldots, S_{n_F}\}$ $\leftarrow$ stack response maps, get new 3D image

  $H = \max(0, S)$ $\leftarrow$ apply ReLu

  $\mathbf{s} = W \text{vec}(H) + \mathbf{b}$ $\leftarrow$ fully-connected layer to get C scores

  $\mathbf{p} = \text{SOFTMAX}(\mathbf{s})$ $\leftarrow$ turn scores into probabilities

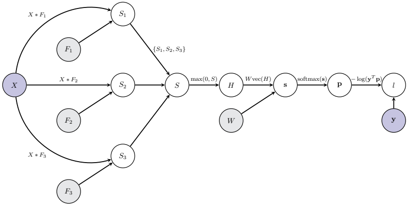
- Dimensions of inputs, outputs and parameters:
  - X is $w \times h \times 3$
  - Each F_i is $f \times f \times 3$ and b_i is a scalar
  - Each S_i is $(w - f + 1) \times (h - f + 1)$
  - S and H are $(w - f + 1) \times (h - f + 1) \times n_F$
  - W is $C \times (w - f + 1)(h - f + 1)n_F$
  - $\mathbf{b}$, $\mathbf{s}$ and $\mathbf{p}$ are $C \times 1$.

How do we learn the parameters of the network?

- Optimize the usual cross-entropy loss $(+\ L_2$ regularization term) on the training data.
- Use mini-batch gradient descent to perform optimization.
- $\implies$ need to compute the gradient of the loss w.r.t. the convolutional parameters....

Gradient Computations for one Convolutional layer

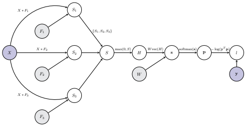
Computational Graph for our simple network



Notes about the above figure

- Apply 3 filters in the convolutional layer ($n_F = 3$).
- $X = \{X_1, X_2, X_3\}$ and each X_i has size $w \times h$
- Each $F_i = \{F_{i1}, F_{i2}, F_{i3}\}$ and has size $f \times f \times 3$
- Have omitted the bias weights for clarity.

Computational Graph for our simple network



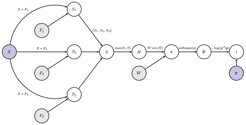
From previous lectures know that

$$\frac{\partial l}{\partial \mathbf{s}} = -\frac{\mathbf{y}^T}{\mathbf{y}^T\mathbf{p}}\left(\text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^T\right)$$

$$\frac{\partial l}{\partial \text{vec}(H)} = \frac{\partial l}{\partial \mathbf{s}} W$$

$$\frac{\partial l}{\partial \text{vec}(S)} = \frac{\partial l}{\partial \text{vec}(H)} \text{diag}\left(\text{Ind}(\text{vec}(S) > 0)\right)$$

Computational Graph for our simple network



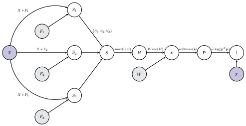
From reading the computational graph we can see that

$$\frac{\partial l}{\partial \text{vec}(F_i)} = \frac{\partial l}{\partial \text{vec}(S_i)} \frac{\partial \text{vec}(S_i)}{\partial \text{vec}(F_i)}$$

$$= \frac{\partial l}{\partial \text{vec}(S)} \frac{\partial \text{vec}(S)}{\partial \text{vec}(S_i)} \frac{\partial \text{vec}(S_i)}{\partial \text{vec}(F_i)}$$

for $i = 1, 2, 3$.

Computational Graph for our simple network

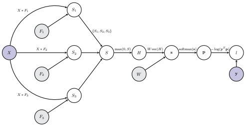


From reading the computational graph we can see that

$$\frac{\partial l}{\partial \text{vec}(F_i)} = \underbrace{\frac{\partial l}{\partial \text{vec}(S)}}_{\uparrow \text{ already know}} \frac{\partial \text{vec}(S)}{\partial \text{vec}(S_i)} \frac{\partial \text{vec}(S_i)}{\partial \text{vec}(F_i)}$$

for $i = 1, 2, 3$.

Computational Graph for our simple network



From reading the computational graph we can see that

$$\frac{\partial l}{\partial \text{vec}(F_i)} = \frac{\partial l}{\partial \text{vec}(S)} \underbrace{\frac{\partial \text{vec}(S)}{\partial \text{vec}(S_i)}}_{\uparrow \text{ calculate now}} \frac{\partial \text{vec}(S_i)}{\partial \text{vec}(F_i)}$$

for $i = 1, 2, 3$.

Jacobian of vec(S) w.r.t. vec(S_i)

- Have $S = \{S_1, S_2, S_3\} \implies$

$$\text{vec}(S) = \begin{pmatrix} \text{vec}(S_1) \\ \text{vec}(S_2) \\ \text{vec}(S_3) \end{pmatrix}$$

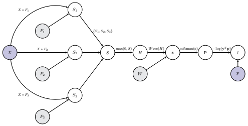
- Then

$$\frac{\partial \text{vec}(S)}{\partial \text{vec}(S_1)} = \begin{pmatrix} I_t \\ 0 \\ 0 \end{pmatrix}, \quad \frac{\partial \text{vec}(S)}{\partial \text{vec}(S_2)} = \begin{pmatrix} 0 \\ I_t \\ 0 \end{pmatrix}, \quad \frac{\partial \text{vec}(S)}{\partial \text{vec}(S_3)} = \begin{pmatrix} 0 \\ 0 \\ I_t \end{pmatrix}$$

  where $t = (w - f + 1) \times (h - f + 1)$ and each 0 denotes a square matrix of zeros of size $t \times t$.

- Each $\frac{\partial \text{vec}(S)}{\partial \text{vec}(S_i)}$ has size $3t \times t$

Computational Graph for our simple network



From reading the computational graph we can see that

$$\frac{\partial l}{\partial \text{vec}(F_i)} = \frac{\partial l}{\partial \text{vec}(S)} \frac{\partial \text{vec}(S)}{\partial \text{vec}(S_i)} \frac{\partial \text{vec}(S_i)}{\partial \text{vec}(F_i)}$$

↑ calculate now

for $i = 1, 2, 3$.

Jacobian of vec(S_i) w.r.t. vec(F_i)

- Have for $i = 1, 2, 3$:

$$S_i = X * F_i$$

- Can write a convolution (not in a very memory efficient way) as a matrix multiplication

$$\text{vec}(S_i) = M_X^{\text{im}} \, \text{vec}(F_i)$$

- M_X^{im} has size $(w - f + 1)(h - f + 1) \times (3f^2)$

- What are the entries of M_X^{im}?

Writing a convolution as a matrix multiplication

Simple Example

- Have an input image X of size $6 \times 6 \times 1$.

- Have a filter F of size $3 \times 3 \times 1$.

- Convolve X by F gives a response map of size 4×4

$$S = X * F$$

- Each entry of S can be written as

$$S_{lm} = \sum_{k=1}^{1} \sum_{i=1}^{3} \sum_{j=1}^{3} X_{k,i+l-1,j+m-1} F_{kij}$$

Writing a convolution as a matrix multiplication

Simple Example

Want to write this convolution as a matrix multiplication:

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

One Solution:

$$S_{11} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}$$

new row corresponds to

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing a convolution as a matrix multiplication

Simple Example

Want to write this convolution as a matrix multiplication:

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

One Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}$$

new row corresponds to

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing a convolution as a matrix multiplication

Simple Example

Want to write this convolution as a matrix multiplication:

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

One Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}$$

new row corresponds to

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing a convolution as a matrix multiplication

Simple Example

Want to write this convolution as a matrix multiplication:

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

One Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}$$

new row corresponds to

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Simple Example

Want to write this convolution as a matrix multiplication:

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

One Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \\ S_{15} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}$$

new row corresponds to $\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$

Simple Example

Want to write this convolution as a matrix multiplication:

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

One Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \\ S_{21} \\ \vdots \\ \vdots \\ S_{44} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & X_{45} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}$$

new row corresponds to $\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$

Simple Example

Want to write this convolution as a matrix multiplication:

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

One Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \\ S_{21} \\ \vdots \\ \vdots \\ S_{44} \end{pmatrix} = \underbrace{\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & X_{45} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}$$

new row corresponds to $\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$

$$\text{vec}(S) = M_X^{\text{im}}\, \text{vec}(F)$$

Multiple planes: Convolution → Matrix multiplication

- What about when X and F_1 have multiple planes?
- Say $X = \{X_1, X_2, X_3, X_4\}$ has size $6 \times 6 \times 4$,
- $F_1 = \{F_{11}, F_{12}, F_{13}, F_{14}\}$ has size $3 \times 3 \times 4$.
- Let

$$S_1 = X * F_1 \qquad (S_1 \text{ has size } 4 \times 4)$$

- Then

$$\text{vec}(S_1) = M_X^{\text{im}}\, \text{vec}(F_1)$$

where

$$M_X^{\text{im}} = \begin{pmatrix} M_{X_1}^{\text{im}} & M_{X_2}^{\text{im}} & M_{X_3}^{\text{im}} & M_{X_4}^{\text{im}} \end{pmatrix}$$

and has size 16×36.

Back to: Jacobian of vec(S_i) w.r.t. vec(F_i)

- Have for $i = 1, 2, 3$:

$$S_i = X * F_i$$

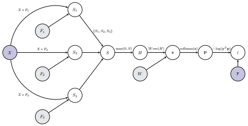
- Can write a convolution (not in a very memory efficient way) as a matrix multiplication

$$\text{vec}(S_i) = M_X^{\text{im}} \text{vec}(F_i)$$

- M_X^{im} has size $(w - f + 1)(h - f + 1) \times (3f^2)$
- Thus

$$\frac{\partial \text{vec}(S_i)}{\partial \text{vec}(F_i)} = M_X^{\text{im}}$$

Gradient of the loss w.r.t. vec(F_i)



Thus

$$\frac{\partial l}{\partial \text{vec}(F_1)} = \frac{\partial l}{\partial \text{vec}(S)} \frac{\partial \text{vec}(S)}{\partial \text{vec}(S_1)} \frac{\partial \text{vec}(S_1)}{\partial \text{vec}(F_1)} = \frac{\partial l}{\partial \text{vec}(S)} \begin{pmatrix} I_t \\ 0 \\ 0 \end{pmatrix} M_X^{\text{im}}$$

$$= \begin{pmatrix} \frac{\partial l}{\partial \text{vec}(S_1)} & \frac{\partial l}{\partial \text{vec}(S_2)} & \frac{\partial l}{\partial \text{vec}(S_3)} \end{pmatrix} \begin{pmatrix} M_{X_1}^{\text{im}} & M_{X_2}^{\text{im}} & M_{X_3}^{\text{im}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \sum_{i=1}^{3} \frac{\partial l}{\partial \text{vec}(S_1)} M_{X_i}^{\text{im}}$$

Gradient of the loss w.r.t. F_i

- May want expression for $\frac{\partial l}{\partial F_i}$ instead of $\frac{\partial l}{\partial \text{vec}(F_i)}$.
- **Option 1**:

  Reshape $\frac{\partial l}{\partial \text{vec}(F_i)}$ (size $1 \times 3f^2$) to $\frac{\partial l}{\partial F_i}$ (size $f \times f \times 3$).

Gradient of the loss w.r.t. F_i

- May want expression for $\frac{\partial l}{\partial F_i}$ instead of $\frac{\partial l}{\partial \text{vec}(F_i)}$.
- **Option 2**:

  Return to our simple example ...

Writing a certain matrix multiplication as a convolution

Return to Simple Example

Consider the case

$$(v_1 \quad v_2 \quad \cdots \quad v_9) = (g_1 \quad g_2 \quad \cdots \quad g_{16}) \underbrace{\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & X_{45} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9}$$

Writing a certain matrix multiplication as a convolution

Return to Simple Example

Consider the case

$$({\color{red}v_1} \quad v_2 \quad \cdots \quad v_9) = (g_1 \quad g_2 \quad \cdots \quad g_{16}) \underbrace{\begin{pmatrix} {\color{red}X_{11}} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ {\color{red}X_{12}} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ {\color{red}X_{13}} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ {\color{red}X_{14}} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ {\color{red}X_{21}} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ {\color{red}X_{44}} & X_{45} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9}$$

where red column in M_X^{im} corresponds to this red block in X

$$\begin{pmatrix} {\color{red}X_{11}} & {\color{red}X_{12}} & {\color{red}X_{13}} & {\color{red}X_{14}} & X_{15} & X_{16} \\ {\color{red}X_{21}} & {\color{red}X_{22}} & {\color{red}X_{23}} & {\color{red}X_{24}} & X_{25} & X_{26} \\ {\color{red}X_{31}} & {\color{red}X_{32}} & {\color{red}X_{33}} & {\color{red}X_{34}} & X_{35} & X_{36} \\ {\color{red}X_{41}} & {\color{red}X_{42}} & {\color{red}X_{43}} & {\color{red}X_{44}} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing a certain matrix multiplication as a convolution

Return to Simple Example

Consider the case

$$(v_1 \quad {\color{red}v_2} \quad \cdots \quad v_9) = (g_1 \quad g_2 \quad \cdots \quad g_{16}) \underbrace{\begin{pmatrix} X_{11} & {\color{red}X_{12}} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & {\color{red}X_{13}} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & {\color{red}X_{14}} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & {\color{red}X_{15}} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & {\color{red}X_{22}} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & {\color{red}X_{45}} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9}$$

where red column in M_X^{im} corresponds to this red block in X

$$\begin{pmatrix} X_{11} & {\color{red}X_{12}} & {\color{red}X_{13}} & {\color{red}X_{14}} & {\color{red}X_{15}} & X_{16} \\ X_{21} & {\color{red}X_{22}} & {\color{red}X_{23}} & {\color{red}X_{24}} & {\color{red}X_{25}} & X_{26} \\ X_{31} & {\color{red}X_{32}} & {\color{red}X_{33}} & {\color{red}X_{34}} & {\color{red}X_{35}} & X_{36} \\ X_{41} & {\color{red}X_{42}} & {\color{red}X_{43}} & {\color{red}X_{44}} & {\color{red}X_{45}} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing a certain matrix multiplication as a convolution

Return to Simple Example

Consider the case

$$(v_1 \quad v_2 \quad {\color{red}v_3} \quad \cdots \quad v_9) = (g_1 \quad g_2 \quad \cdots \quad g_{16}) \underbrace{\begin{pmatrix} X_{11} & X_{12} & {\color{red}X_{13}} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & {\color{red}X_{14}} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & {\color{red}X_{15}} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & {\color{red}X_{16}} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & {\color{red}X_{23}} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & X_{45} & {\color{red}X_{46}} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9}$$

where red column in M_X^{im} corresponds to this red block in X

$$\begin{pmatrix} X_{11} & X_{12} & {\color{red}X_{13}} & {\color{red}X_{14}} & {\color{red}X_{15}} & {\color{red}X_{16}} \\ X_{21} & X_{22} & {\color{red}X_{23}} & {\color{red}X_{24}} & {\color{red}X_{25}} & {\color{red}X_{26}} \\ X_{31} & X_{32} & {\color{red}X_{33}} & {\color{red}X_{34}} & {\color{red}X_{35}} & {\color{red}X_{36}} \\ X_{41} & X_{42} & {\color{red}X_{43}} & {\color{red}X_{44}} & {\color{red}X_{45}} & {\color{red}X_{46}} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Return to Simple Example

Consider the case

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & \cdots & v_9 \end{pmatrix} = \begin{pmatrix} g_1 & g_2 & \cdots & g_{16} \end{pmatrix} \underbrace{\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & X_{45} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9}$$

where red column in M_X^{im} corresponds to this red block in X

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Return to Simple Example

$$\cdots$$

Return to Simple Example

Consider the case

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & \cdots & v_9 \end{pmatrix} = \begin{pmatrix} g_1 & g_2 & \cdots & g_{16} \end{pmatrix} \underbrace{\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & X_{45} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9}$$

where red column in M_X^{im} corresponds to this red block in X

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Return to Simple Example

Consider the case

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & \cdots & v_9 \end{pmatrix} = \begin{pmatrix} g_1 & g_2 & \cdots & g_{16} \end{pmatrix} \underbrace{\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} \\ X_{12} & X_{13} & X_{14} & X_{22} & X_{23} & X_{24} & X_{32} & X_{33} & X_{34} \\ X_{13} & X_{14} & X_{15} & X_{23} & X_{24} & X_{25} & X_{33} & X_{34} & X_{35} \\ X_{14} & X_{15} & X_{16} & X_{24} & X_{25} & X_{26} & X_{34} & X_{35} & X_{36} \\ X_{21} & X_{22} & X_{23} & X_{31} & X_{32} & X_{33} & X_{41} & X_{42} & X_{43} \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \\ X_{44} & X_{45} & X_{46} & X_{54} & X_{55} & X_{56} & X_{64} & X_{65} & X_{66} \end{pmatrix}}_{M_X^{\text{im}} \text{ size } 16 \times 9}$$

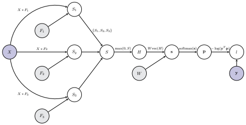
where red column in M_X^{im} corresponds to this red block in X

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Thus

$$\begin{pmatrix} v_1 & v_2 & v_3 \\ v_4 & v_5 & v_6 \\ v_7 & v_8 & v_9 \end{pmatrix} = X * \begin{pmatrix} g_1 & g_2 & g_3 & g_4 \\ g_5 & g_6 & g_7 & g_8 \\ g_9 & g_{10} & g_{11} & g_{12} \\ g_{13} & g_{14} & g_{15} & g_{16} \end{pmatrix}$$

Back to Gradient of the loss w.r.t. F_i



Know

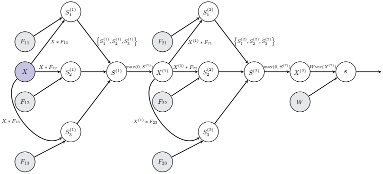
$$\frac{\partial l}{\partial \text{vec}(F_i)} = \sum_{j=1}^{3} \frac{\partial l}{\partial \text{vec}(S_i)} M^{\text{im}}_{X_j}$$

but our simple example $\Longrightarrow$

$$\frac{\partial l}{\partial F_i} = \sum_{j=1}^{3} X_j * \frac{\partial l}{\partial S_i}$$

Gradient Computations for two Convolutional layers

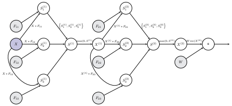
Computational Graph: two convolutional layers



Notes about the figure

- Apply 3 filters at each convolutional layer.
- Have omitted the bias weights for clarity.

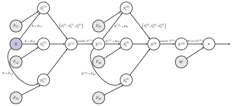
How do we back-propagate the gradient to node $X^{(1)}$?



- Children of node $X^{(1)}$ are $S_1^{(2)}, S_2^{(2)}$ and $S_3^{(2)}$
- Thus

$$\frac{\partial l}{\partial \text{vec}(X^{(1)})} = \sum_{i=1}^{3} \frac{\partial l}{\partial \text{vec}(S_i^{(2)})} \frac{\partial \text{vec}(S_i^{(2)})}{\partial \text{vec}(X^{(1)})}$$

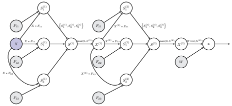
How do we back-propagate the gradient to node $X^{(1)}$?



- Children of node $X^{(1)}$ are $S_1^{(2)}, S_2^{(2)}$ and $S_3^{(2)}$
- Thus

$$\frac{\partial l}{\partial \text{vec}(X^{(1)})} = \sum_{i=1}^{3} \underbrace{\frac{\partial l}{\partial \text{vec}(S_i^{(2)})}}_{\text{already know}} \frac{\partial \text{vec}(S_i^{(2)})}{\partial \text{vec}(X^{(1)})}$$

How do we back-propagate the gradient to node $X^{(1)}$?



- Children of node $X^{(1)}$ are $S_1^{(2)}, S_2^{(2)}$ and $S_3^{(2)}$
- Thus

$$\frac{\partial l}{\partial \text{vec}(X^{(1)})} = \sum_{i=1}^{3} \frac{\partial l}{\partial \text{vec}(S_i^{(2)})} \underbrace{\frac{\partial \text{vec}(S_i^{(2)})}{\partial \text{vec}(X^{(1)})}}_{\text{calculate now}}$$

Jacobian of vec$(S_i^{(2)})$ w.r.t. vec$(X^{(1)})$

- Have for $i = 1, 2, 3$:

$$S_i^{(2)} = X^{(1)} * F_{2i}$$

- Can write a convolution (not in a very memory efficient way) as a matrix multiplication

$$\text{vec}(S_i^{(2)}) = M_{F_{2i}}^{\text{filter}} \, \text{vec}(X^{(1)})$$

- $M_{F_{2i}}$ has size $(w - f + 1)(h - f + 1) \times 3wh$ (assuming $X^{(1)}$ has size $w \times h \times 3$ and F_{2i} has size $f \times f \times 3$.)
- What are the entries of $M_{F_{2i}}^{\text{filter}}$?

Writing convolution as a matrix multiplication

Simple Example

- Have an input image X of size $6 \times 6 \times 1$.
- Have a filter F of size $3 \times 3 \times 1$.
- Convolve X by F gives a response map of size 4×4

$$S = X * F$$

- Each entry of S can be written as

$$S_{lm} = \sum_{k=1}^{1} \sum_{i=1}^{3} \sum_{j=1}^{3} X_{k,i+l-1,j+m-1} F_{kij}$$

Writing convolution as a matrix multiplication

Simple Example

Write this convolution as a matrix multiplication involving vec(X)

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

Solution:

$$S_{11} = (\overbrace{F_{11}\ F_{12}\ F_{13}\ 0\ 0\ 0}^{\text{entries corresponding to row 1 of } X}\ \overbrace{F_{21}\ F_{22}\ F_{23}\ 0\ 0\ 0}^{\text{row 2 of } X}\ \overbrace{F_{31}\ F_{32}\ F_{33}\ 0\ 0\ 0}^{\text{row 3 of } X}\ \cdots\ )\,\text{vec}(X)$$

S_{11} is the dot product between F and red entries of X:

$$\begin{pmatrix} \color{red}{X_{11}} & \color{red}{X_{12}} & \color{red}{X_{13}} & X_{14} & X_{15} & X_{16} \\ \color{red}{X_{21}} & \color{red}{X_{22}} & \color{red}{X_{23}} & X_{24} & X_{25} & X_{26} \\ \color{red}{X_{31}} & \color{red}{X_{32}} & \color{red}{X_{33}} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing convolution as a matrix multiplication

Simple Example

Write this convolution as a matrix multiplication involving vec(X)

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \end{pmatrix} = \begin{pmatrix} \overbrace{F_{11} & F_{12} & F_{13} & 0 & 0 & 0}^{\text{entries corresponding to row 1 of } X} & \overbrace{F_{21} & F_{22} & F_{23} & 0 & 0 & 0}^{\text{row 2 of } X} & \overbrace{F_{31} & F_{32} & F_{33} & 0 & 0 & 0}^{\text{row 3 of } X} & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \end{pmatrix} \text{vec}(X)$$

S_{12} is the dot product between F and red entries of X:

$$\begin{pmatrix} X_{11} & \color{red}{X_{12}} & \color{red}{X_{13}} & \color{red}{X_{14}} & X_{15} & X_{16} \\ X_{21} & \color{red}{X_{22}} & \color{red}{X_{23}} & \color{red}{X_{24}} & X_{25} & X_{26} \\ X_{31} & \color{red}{X_{32}} & \color{red}{X_{33}} & \color{red}{X_{34}} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing convolution as a matrix multiplication

Simple Example

Write this convolution as a matrix multiplication involving vec(X)

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \end{pmatrix} = \begin{pmatrix} \overbrace{F_{11} & F_{12} & F_{13} & 0 & 0 & 0}^{\text{entries corresponding to row 1 of } X} & \overbrace{F_{21} & F_{22} & F_{23} & 0 & 0 & 0}^{\text{row 2 of } X} & \overbrace{F_{31} & F_{32} & F_{33} & 0 & 0 & 0}^{\text{row 3 of } X} & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \end{pmatrix} \text{vec}(X)$$

S_{13} is the dot product between F and red entries of X:

$$\begin{pmatrix} X_{11} & X_{12} & \color{red}{X_{13}} & \color{red}{X_{14}} & \color{red}{X_{15}} & X_{16} \\ X_{21} & X_{22} & \color{red}{X_{23}} & \color{red}{X_{24}} & \color{red}{X_{25}} & X_{26} \\ X_{31} & X_{32} & \color{red}{X_{33}} & \color{red}{X_{34}} & \color{red}{X_{35}} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing convolution as a matrix multiplication

Simple Example

Write this convolution as a matrix multiplication involving vec(X)

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \end{pmatrix} = \begin{pmatrix} \overbrace{F_{11} & F_{12} & F_{13} & 0 & 0 & 0}^{\text{entries corresponding to row 1 of } X} & \overbrace{F_{21} & F_{22} & F_{23} & 0 & 0 & 0}^{\text{row 2 of } X} & \overbrace{F_{31} & F_{32} & F_{33} & 0 & 0 & 0}^{\text{row 3 of } X} & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \end{pmatrix} \text{vec}(X)$$

S_{14} is the dot product between F and red entries of X:

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & \color{red}{X_{14}} & \color{red}{X_{15}} & \color{red}{X_{16}} \\ X_{21} & X_{22} & X_{23} & \color{red}{X_{24}} & \color{red}{X_{25}} & \color{red}{X_{26}} \\ X_{31} & X_{32} & X_{33} & \color{red}{X_{34}} & \color{red}{X_{35}} & \color{red}{X_{36}} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing convolution as a matrix multiplication

Simple Example

Write this convolution as a matrix multiplication involving vec(X)

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \\ S_{21} \end{pmatrix} = \begin{pmatrix} \overbrace{F_{11} & F_{12} & F_{13} & 0 & 0 & 0}^{\text{entries corresponding to row 1 of } X} & \overbrace{F_{21} & F_{22} & F_{23} & 0 & 0 & 0}^{\text{row 2 of } X} & \overbrace{F_{31} & F_{32} & F_{33} & 0 & 0 & 0}^{\text{row 3 of } X} & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \end{pmatrix} \text{vec}(X)$$

S_{21} is the dot product between F and red entries of X:

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ \color{red}{X_{21}} & \color{red}{X_{22}} & \color{red}{X_{23}} & X_{24} & X_{25} & X_{26} \\ \color{red}{X_{31}} & \color{red}{X_{32}} & \color{red}{X_{33}} & X_{34} & X_{35} & X_{36} \\ \color{red}{X_{41}} & \color{red}{X_{42}} & \color{red}{X_{43}} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix}$$

Writing convolution as a matrix multiplication

Simple Example

Write this convolution as a matrix multiplication involving vec(X)

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \\ S_{21} \\ \vdots \\ \vdots \end{pmatrix} = \begin{pmatrix} \overbrace{F_{11} & F_{12} & F_{13} & 0 & 0 & 0}^{\text{entries corresponding to row 1 of } X} & \overbrace{F_{21} & F_{22} & F_{23} & 0 & 0 & 0}^{\text{row 2 of } X} & \overbrace{F_{31} & F_{32} & F_{33} & 0 & 0 & 0}^{\text{row 3 of } X} & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \text{vec}(X)$$

Writing convolution as a matrix multiplication

Simple Example

Write this convolution as a matrix multiplication involving vec(X)

$$S = \begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} & X_{16} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} & X_{26} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} & X_{36} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} & X_{46} \\ X_{51} & X_{52} & X_{53} & X_{54} & X_{55} & X_{56} \\ X_{61} & X_{62} & X_{63} & X_{64} & X_{65} & X_{66} \end{pmatrix} * \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}$$

Solution:

$$\begin{pmatrix} S_{11} \\ S_{12} \\ S_{13} \\ S_{14} \\ S_{21} \\ \vdots \\ \vdots \end{pmatrix} = \underbrace{\begin{pmatrix} \overbrace{F_{11} & F_{12} & F_{13} & 0 & 0 & 0}^{\text{entries corresponding to row 1 of } X} & \overbrace{F_{21} & F_{22} & F_{23} & 0 & 0 & 0}^{\text{row 2 of } X} & \overbrace{F_{31} & F_{32} & F_{33} & 0 & 0 & 0}^{\text{row 3 of } X} & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}}_{M_F^{\text{filter}}} \text{vec}(X)$$

Thus

$$\text{vec}(S) = M_F^{\text{filter}}\, \text{vec}(X)$$

Multiple planes: Convolution → Matrix multiplication

- What about when X and F_1 have multiple planes?
- $X = \{X_1, X_2, X_3, X_4\}$ has size $6 \times 6 \times 4$
- $F_1 = \{F_{11}, F_{12}, F_{13}, F_{14}\}$ has size $3 \times 3 \times 4$
- Let

$$S_1 = X * F_1$$

- Then

$$\text{vec}(S_1) = M_{F_1}^{\text{filter}} \text{vec}(X)$$

where

$$M_{F_1}^{\text{filter}} = \begin{pmatrix} M_{F_{11}}^{\text{filter}} & M_{F_{12}}^{\text{filter}} & M_{F_{13}}^{\text{filter}} & M_{F_{14}}^{\text{filter}} \end{pmatrix}$$

and has size 16×144.

Back to: Jacobian of vec($S_i^{(2)}$) w.r.t. vec($X^{(1)}$)

- Have for $i = 1, 2, 3$:

$$S_i^{(2)} = X^{(1)} * F_{2i}$$

- Can write a convolution (not in a very memory efficient way) as a matrix multiplication

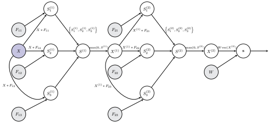
$$\text{vec}(S_i^{(2)}) = M_{F_{2i}}^{\text{filter}} \text{vec}(X^{(1)})$$

- $M_{F_{2i}}^{\text{filter}}$ has size $(w' - f + 1)(h' - f + 1) \times 3w'h'$ (where $w' = w - f + 1$ and $h' = h - f + 1$).

- Thus

$$\frac{\partial \text{vec}(S_i^{(2)})}{\partial \text{vec}(X^{(1)})} = M_{F_{2i}}^{\text{filter}}$$

Gradient of the loss w.r.t. node vec($X^{(1)}$)



- Thus

$$\frac{\partial l}{\partial \text{vec}(X^{(1)})} = \sum_{i=1}^{3} \frac{\partial l}{\partial \text{vec}(S_i^{(2)})} \frac{\partial \text{vec}(S_i^{(2)})}{\partial \text{vec}(X^{(1)})}$$

$$= \sum_{i=1}^{3} \frac{\partial l}{\partial \text{vec}(S_i^{(2)})} M_{F_{2i}}^{\text{filter}}$$

Gradient of the loss w.r.t. $X^{(1)}$

- May want expression for $\frac{\partial l}{\partial X^{(1)}}$ instead of $\frac{\partial l}{\partial \text{vec}(X^{(1)})}$.
- **Option 1**:

  Reshape $\frac{\partial l}{\partial \text{vec}(X^{(1)})}$ (size $1 \times 3w'h'$) to $\frac{\partial l}{\partial X^{(1)}}$ (size $w' \times h' \times 3$).

Gradient of the loss w.r.t. $X^{(1)}$

- May want expression for $\frac{\partial l}{\partial X^{(1)}}$ instead of $\frac{\partial l}{\partial \text{vec}(X^{(1)})}$.
- **Option 2**:

  Return to our simple example ...

Turn our matrix multiplication to convolution

Consider the case

$$(v_1 \quad v_2 \quad \dots \quad v_{36}) = (g_1 \quad g_2 \quad \dots \quad g_{16}) \begin{pmatrix} F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & 0 & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Turn matrix multiplcation to convolution

$$(v_1 \quad v_2 \quad \dots \quad v_{36}) = (g_1 \quad g_2 \quad \dots \quad g_{16}) \begin{pmatrix} F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & 0 & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$v_1 = g_1 F_{11}$

Turn matrix multiplcation to convolution

$$(v_1 \quad v_2 \quad \dots \quad v_{36}) = (g_1 \quad g_2 \quad \dots \quad g_{16}) \begin{pmatrix} F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & 0 & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$v_1 = g_1 F_{11}$

$v_2 = g_1 F_{12} + g_2 F_{11}$

Turn matrix multiplcation to convolution

$$(v_1 \quad v_2 \quad \dots \quad v_{36}) = (g_1 \quad g_2 \quad \dots \quad g_{16}) \begin{pmatrix} F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & 0 & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$v_1 = g_1 F_{11}$

$v_2 = g_1 F_{12} + g_2 F_{11}$

$v_3 = g_1 F_{13} + g_2 F_{12} + g_3 F_{11}$

Turn matrix multiplcation to convolution

$$(v_1 \;\; v_2 \;\; \cdots \;\; v_{36}) = (g_1 \;\; g_2 \;\; \cdots \;\; g_{16}) \begin{pmatrix} F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & 0 & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$\begin{aligned} v_1 &= g_1 F_{11} \\ v_2 &= g_1 F_{12} + g_2 F_{11} \\ v_3 &= g_1 F_{13} + g_2 F_{12} + g_3 F_{11} \\ v_4 &= g_2 F_{13} + g_3 F_{12} + g_4 F_{11} \end{aligned}$$

Turn matrix multiplcation to convolution

$$(v_1 \;\; v_2 \;\; \cdots \;\; v_{36}) = (g_1 \;\; g_2 \;\; \cdots \;\; g_{16}) \begin{pmatrix} F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & 0 & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$\begin{aligned} v_1 &= g_1 F_{11} \\ v_2 &= g_1 F_{12} + g_2 F_{11} \\ v_3 &= g_1 F_{13} + g_2 F_{12} + g_3 F_{11} \\ v_4 &= g_2 F_{13} + g_3 F_{12} + g_4 F_{11} \\ v_5 &= g_3 F_{13} + g_4 F_{12} \\ v_6 &= g_4 F_{13} \\ v_7 &= g_1 F_{21} + g_5 F_{11} \\ v_8 &= g_1 F_{22} + g_2 F_{21} + g_5 F_{12} + g_6 F_{11} \\ v_9 &= g_1 F_{23} + g_2 F_{22} + g_3 F_{21} + g_5 F_{13} + g_6 F_{12} + g_7 F_{11} \\ v_{10} &= g_2 F_{23} + g_3 F_{22} + g_4 F_{21} + g_6 F_{13} + g_7 F_{12} + g_8 F_{11} \\ v_{11} &= g_3 F_{23} + g_4 F_{22} + g_7 F_{13} + g_8 F_{12} \\ v_{12} &= g_4 F_{23} + g_8 F_{13} \\ v_{13} &= g_1 F_{31} + g_5 F_{21} + g_9 F_{11} \\ v_{14} &= g_1 F_{32} + g_2 F_{31} + g_5 F_{22} + g_6 F_{21} + g_9 F_{12} + g_{10} F_{11} \\ v_{15} &= g_1 F_{33} + g_2 F_{32} + g_3 F_{31} + g_5 F_{23} + g_6 F_{22} + g_7 F_{21} + g_9 F_{13} + g_{10} F_{12} + g_{11} F_{11} \\ &\vdots \end{aligned}$$

Turn matrix multiplcation to convolution

$$(v_1 \;\; v_2 \;\; \cdots \;\; v_{36}) = (g_1 \;\; g_2 \;\; \cdots \;\; g_{16}) \begin{pmatrix} F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & 0 & \cdots \\ 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & 0 & \cdots \\ 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & 0 & \cdots \\ 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & F_{31} & F_{32} & F_{33} & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & F_{11} & F_{12} & F_{13} & 0 & 0 & 0 & F_{21} & F_{22} & F_{23} & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$\begin{aligned} v_1 &= g_1 F_{11} \\ v_2 &= g_1 F_{12} + g_2 F_{11} \\ v_3 &= g_1 F_{13} + g_2 F_{12} + g_3 F_{11} \\ v_4 &= g_2 F_{13} + g_3 F_{12} + g_4 F_{11} \\ v_5 &= g_3 F_{13} + g_4 F_{12} \\ v_6 &= g_4 F_{13} \\ v_7 &= g_1 F_{21} + g_5 F_{11} \\ v_8 &= g_1 F_{22} + g_2 F_{21} + g_5 F_{12} + g_6 F_{11} \\ v_9 &= g_1 F_{23} + g_2 F_{22} + g_3 F_{21} + g_5 F_{13} + g_6 F_{12} + g_7 F_{11} \\ v_{10} &= g_2 F_{23} + g_3 F_{22} + g_4 F_{21} + g_6 F_{13} + g_7 F_{12} + g_8 F_{11} \\ v_{11} &= g_3 F_{23} + g_4 F_{22} + g_7 F_{13} + g_8 F_{12} \\ v_{12} &= g_4 F_{23} + g_8 F_{13} \\ v_{13} &= g_1 F_{31} + g_5 F_{21} + g_9 F_{11} \\ v_{14} &= g_1 F_{32} + g_2 F_{31} + g_5 F_{22} + g_6 F_{21} + g_9 F_{12} + g_{10} F_{11} \\ v_{15} &= g_1 F_{33} + g_2 F_{32} + g_3 F_{31} + g_5 F_{23} + g_6 F_{22} + g_7 F_{21} + g_9 F_{13} + g_{10} F_{12} + g_{11} F_{11} \\ &\vdots \end{aligned}$$

There is a pattern here!

Matrix multiplcation → convolution

- Reshape vectors $\mathbf{g}$ and $\mathbf{v}$ into matrices

$$G = \begin{pmatrix} g_1 & g_2 & g_3 & g_4 \\ g_5 & g_6 & g_7 & g_8 \\ g_9 & g_{10} & g_{11} & g_{12} \\ g_{13} & g_{14} & g_{15} & g_{16} \end{pmatrix}, \quad V = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ v_7 & v_8 & v_9 & v_{10} & v_{11} & v_{12} \\ v_{13} & v_{14} & v_{15} & v_{16} & v_{17} & v_{18} \\ v_{19} & v_{20} & v_{21} & v_{22} & v_{23} & v_{24} \\ v_{25} & v_{26} & v_{27} & v_{28} & v_{29} & v_{30} \\ v_{31} & v_{32} & v_{33} & v_{34} & v_{35} & v_{36} \end{pmatrix}$$

- Let

$$G_{\text{zero-pad}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_1 & g_2 & g_3 & g_4 & 0 & 0 \\ 0 & 0 & g_5 & g_6 & g_7 & g_8 & 0 & 0 \\ 0 & 0 & g_9 & g_{10} & g_{11} & g_{12} & 0 & 0 \\ 0 & 0 & g_{13} & g_{14} & g_{15} & g_{16} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad F^{\text{rot180}} = \begin{pmatrix} F_{33} & F_{32} & F_{31} \\ F_{23} & F_{22} & F_{21} \\ F_{13} & F_{12} & F_{11} \end{pmatrix}$$

- Then

$$V = G_{\text{zero-pad}} * F^{\text{rot180}}$$

Multiple planes: Matrix multiplication → Convolution

- $X = \{X_1, X_2, X_3, X_4\}$ has size $6 \times 6 \times 4$
- $F = \{F_1, F_2, F_3, F_4\}$ has size $3 \times 3 \times 4$
- Let

$$S = X * F \qquad (S_1 \text{ has size } 4 \times 4)$$

- Then

$$\text{vec}(S) = M_F^{\text{filter}} \text{vec}(X)$$

where M_F^{filter} has size 16×144.

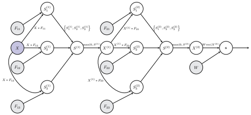
- Let

$$\mathbf{v}^T = \mathbf{g}^T M_F^{\text{filter}}$$

- Let $\text{vec}(V) = \mathbf{v}$ and $\text{vec}(G) = \mathbf{g}$ then

$V = \{G_{\text{zero-pad}} * F_1^{\text{rot180}}, G_{\text{zero-pad}} * F_2^{\text{rot180}}, G_{\text{zero-pad}} * F_3^{\text{rot180}}, G_{\text{zero-pad}} * F_4^{\text{rot180}}\}$

Back to Gradient of the loss w.r.t. node $X^{(1)}$



Know

$$\frac{\partial l}{\partial \text{vec}(X^{(1)})} = \sum_{i=1}^{3} \frac{\partial l}{\partial \text{vec}(S_i^{(2)})} M_{F_{2i}}^{\text{filter}}$$

then

$$\frac{\partial l}{\partial X^{(1)}} = \sum_{i=1}^{3} \left\{ G_i^{\text{zero-pad}} * F_{2i,1}^{\text{rot180}}, G_i^{\text{zero-pad}} * F_{2i,2}^{\text{rot180}}, G_i^{\text{zero-pad}} * F_{2i,3}^{\text{rot180}} \right\}$$

where $G_i = \frac{\partial l}{\partial S_i^{(2)}}$ and $F_{2i} = \{F_{2i,1}, F_{2i,2}, F_{2i,3}\}$.