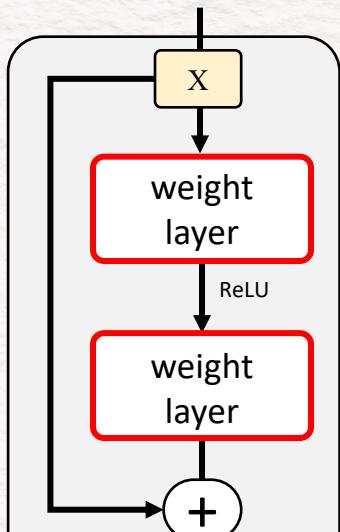


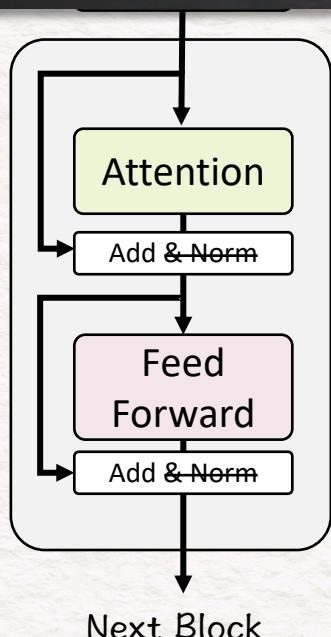
Residual Network



AI by Hand

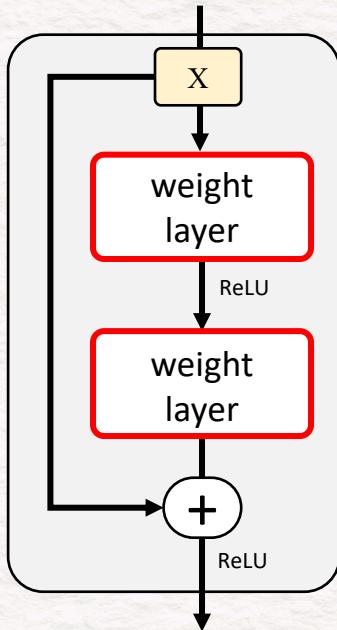


Advanced Series



| | | | | | | | | | | | | | | | | | | | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|---|---|---|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|----|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|---|---|---|---|
| Q K | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td>1</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> </table> | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td></tr> </table> | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 1 | 0 | 1 | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 0 | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0 | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 0 | | | | | | | | | | | | | | | | | | |
| 0 | 0 | 1 | | | | | | | | | | | | | | | | | | |
| <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td>2</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>3</td><td>-2</td></tr> </table> | 2 | 0 | 1 | 1 | 3 | -2 | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table> | | | | | | | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table> | | | | | | |
| 2 | 0 | 1 | | | | | | | | | | | | | | | | | | |
| 1 | 3 | -2 | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> </table> | 1 | 0 | 0 | 0 | 1 | 1 | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>-1</td><td>2</td></tr> </table> | 1 | 1 | 1 | 1 | -1 | 2 | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table> | | | | | | |
| 1 | 0 | 0 | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | | | | | | | | | | | | | | | | | | |
| 1 | -1 | 2 | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td>0</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>0</td></tr> </table> | 0 | 1 | 1 | 1 | 0 | 0 | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td>0</td><td>-1</td><td>3</td></tr> <tr><td>-1</td><td>0</td><td>0</td></tr> </table> | 0 | -1 | 3 | -1 | 0 | 0 | <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table> | | | | | | |
| 0 | 1 | 1 | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0 | | | | | | | | | | | | | | | | | | |
| 0 | -1 | 3 | | | | | | | | | | | | | | | | | | |
| -1 | 0 | 0 | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

Residual Network



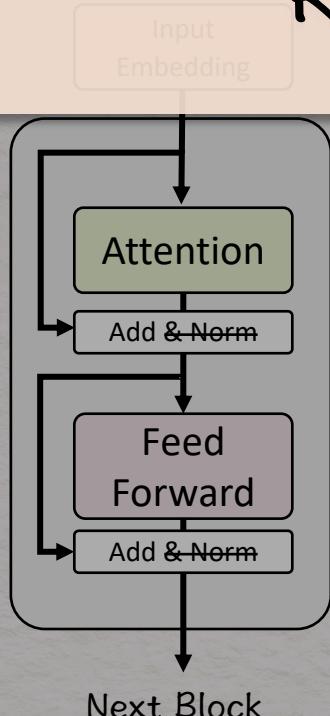
| | | | |
|---|---|---|---|
| | 2 | 1 | 0 |
| | 0 | 3 | 4 |
| | 1 | 0 | 2 |
| 1 | 1 | 1 | 1 |

| | | | |
|---|---|----|----|
| 1 | 1 | 0 | 0 |
| 0 | 1 | -1 | 0 |
| 1 | 0 | 1 | -1 |
| 1 | 1 | 1 | 1 |

| | | | |
|---|----|---|---|
| 1 | 0 | 0 | |
| 0 | 1 | 0 | |
| 0 | 0 | 1 | |
| 0 | 0 | 0 | |
| 0 | -1 | 0 | 0 |
| 1 | 1 | 1 | 1 |

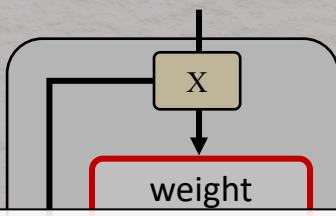
Transformer's Encoder Block

Residual Block



| | | | |
|---|----|----|----|
| Q | 1 | 0 | 1 |
| K | 1 | 1 | 0 |
| | 0 | 1 | 1 |
| | 2 | 0 | 1 |
| | 1 | 3 | -2 |
| | 1 | 1 | 1 |
| | 1 | -1 | 2 |
| | 1 | 0 | 0 |
| | 0 | -1 | 3 |
| | -1 | 0 | 0 |
| 1 | 1 | 1 | 1 |

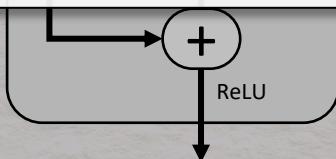
Residual Network



| | | |
|---|---|---|
| 2 | 1 | 0 |
| 0 | 3 | 4 |
| 1 | 0 | 2 |

Given

- A min-batch of 3 input vectors (columns)



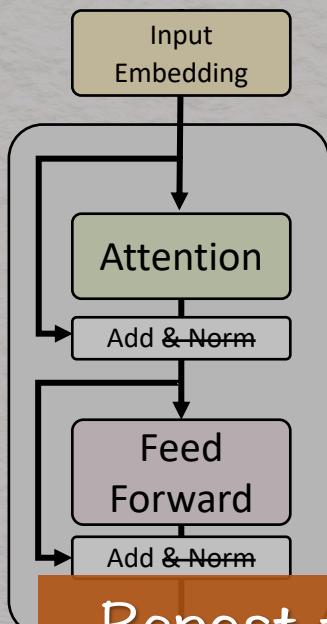
| | | | |
|---|---|----|---|
| 1 | 1 | 0 | 0 |
| 0 | 1 | -1 | 0 |
| 1 | 1 | 1 | 1 |

| | | |
|---|---|---|
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 1 |

| | | | |
|----|----|---|---|
| 0 | 0 | 1 | 0 |
| -1 | 0 | 0 | 0 |
| 0 | -1 | 0 | 0 |

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

Transformer's Encoder Block



| | | | |
|---|---|---|---|
| Q | 1 | 0 | 1 |
| K | 1 | 1 | 0 |
| 0 | 1 | 1 | 1 |

| | | |
|---|---|----|
| 2 | 0 | 1 |
| 1 | 3 | -2 |

| | | |
|---|---|---|
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 1 |

| | | |
|---|---|---|
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 1 |

| | | |
|---|---|---|
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 1 |

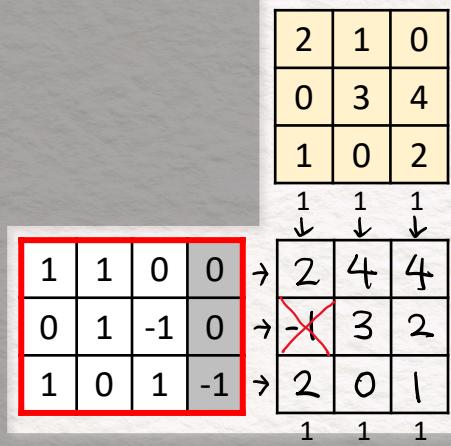
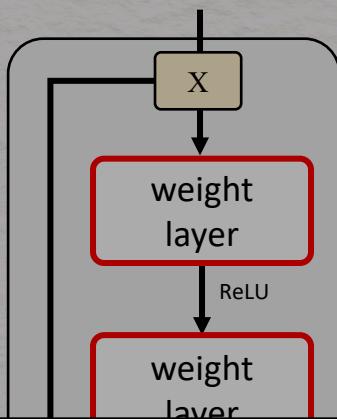
Repost to help other AI learners!

REPOST

| | | |
|----|----|---|
| 1 | 0 | 0 |
| 0 | -1 | 3 |
| -1 | 0 | 0 |

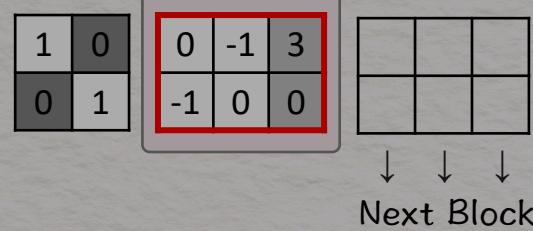
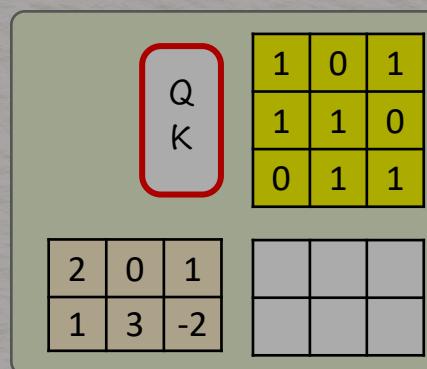
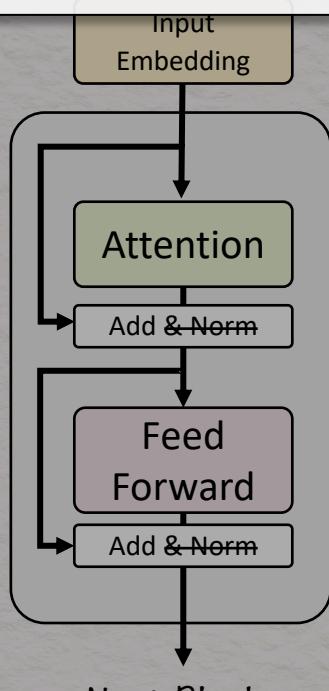
Next Block

Residual Network

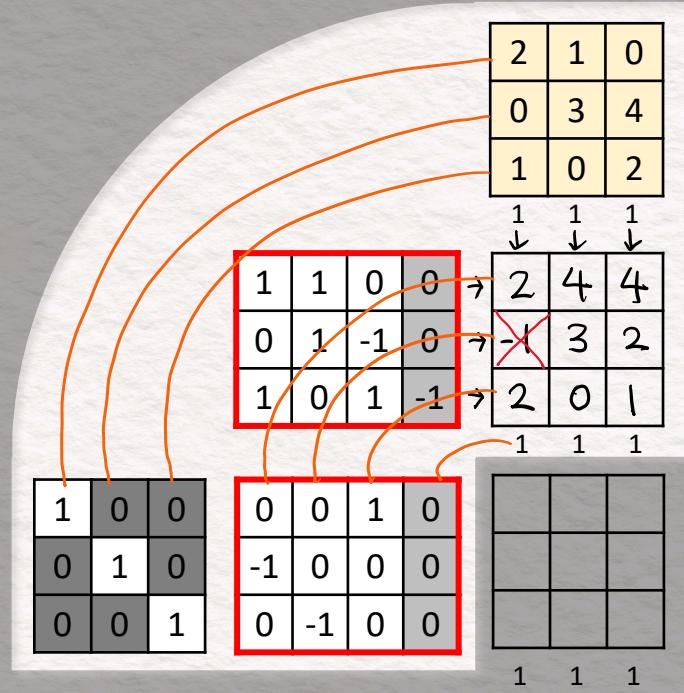
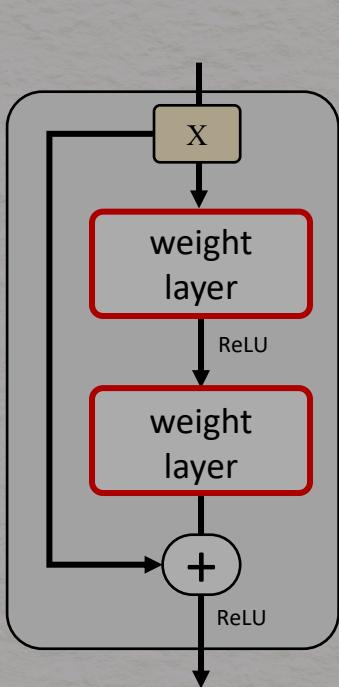


Linear Layer

- Multiply the input vectors with the weight and bias matrix
 - Apply ReLU (negatives \rightarrow 0)
 - Obtain 3 new feature vectors

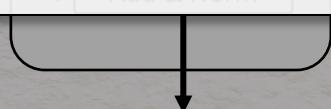


Residual Network

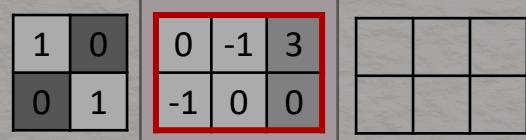


Concatenate

- Stack an identity matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and the weight and bias matrix $\begin{bmatrix} 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}$ of the 2nd layer
- Stack the input vectors $\begin{bmatrix} 2 & 4 & 4 \\ 3 & 2 & 2 \\ 0 & 1 & 1 \end{bmatrix}$ and the feature vectors $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 4 \\ 1 & 0 & 2 \end{bmatrix}$ from the 1st layer
- Draw lines to visualize the links between weights (left) and feature dimensions (top)

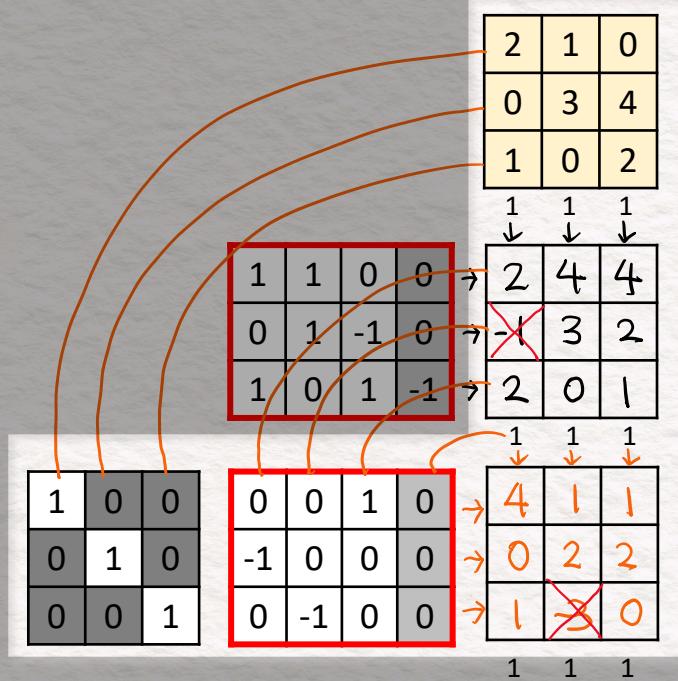
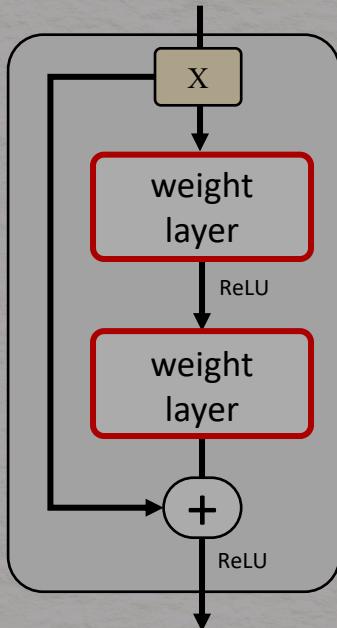


Next Block



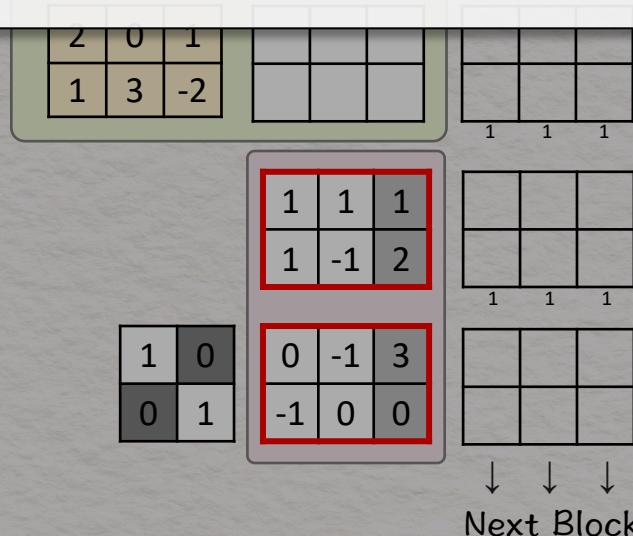
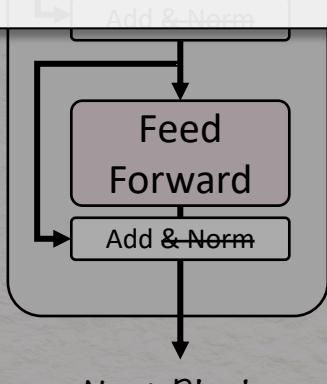
↓ ↓ ↓
Next Block

Residual Network

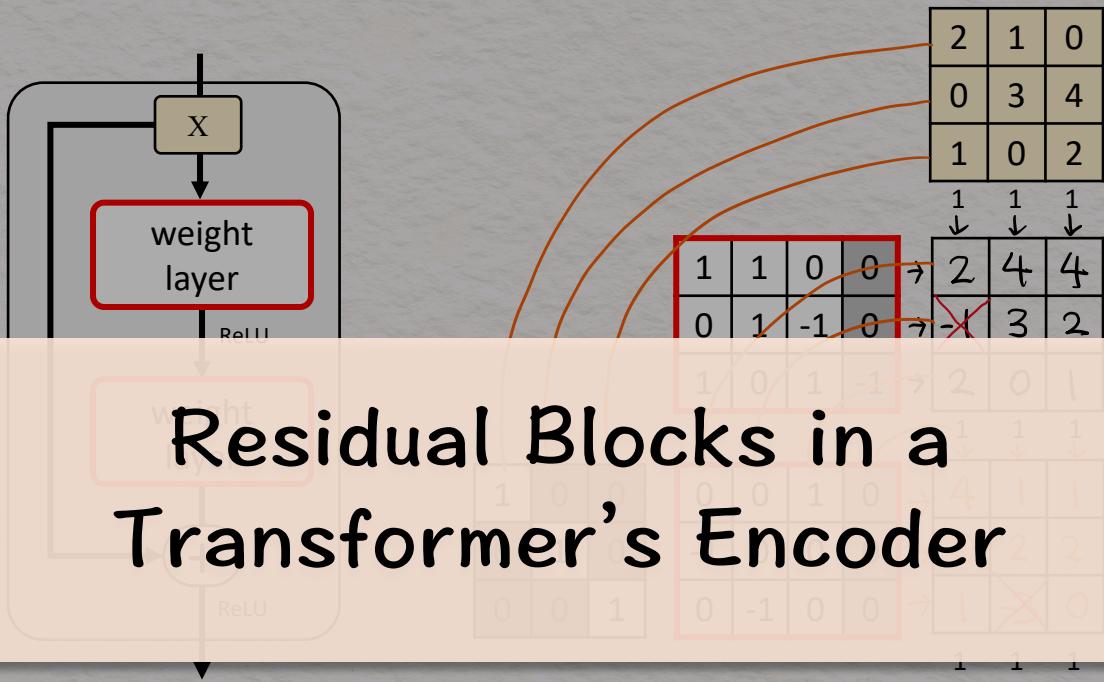


Linear Layer + Identity

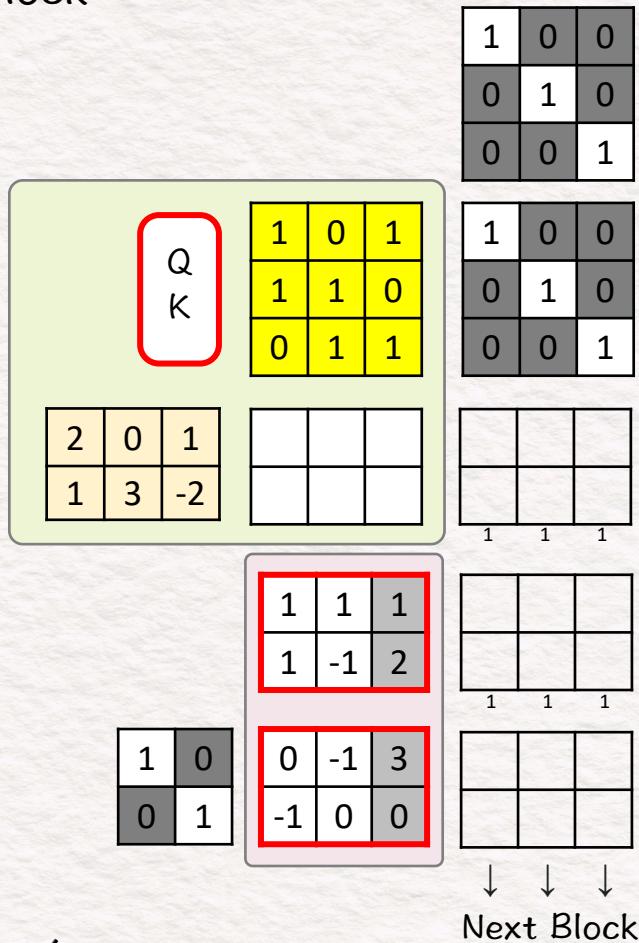
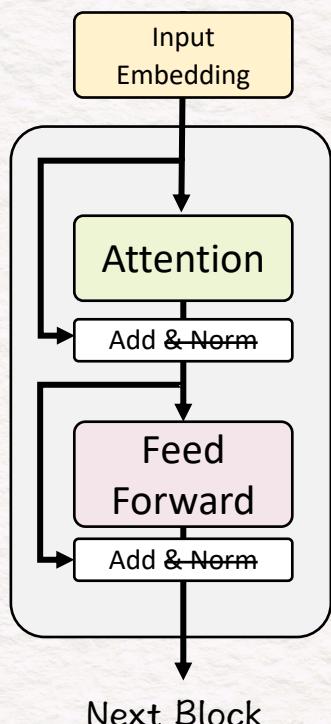
- Multiply the two stacked matrices
- This is equivalent to $F(X) + X$
- Apply ReLU (negatives $\rightarrow 0$)
- Obtain 3 new feature vectors
- Pass them to the next residual block



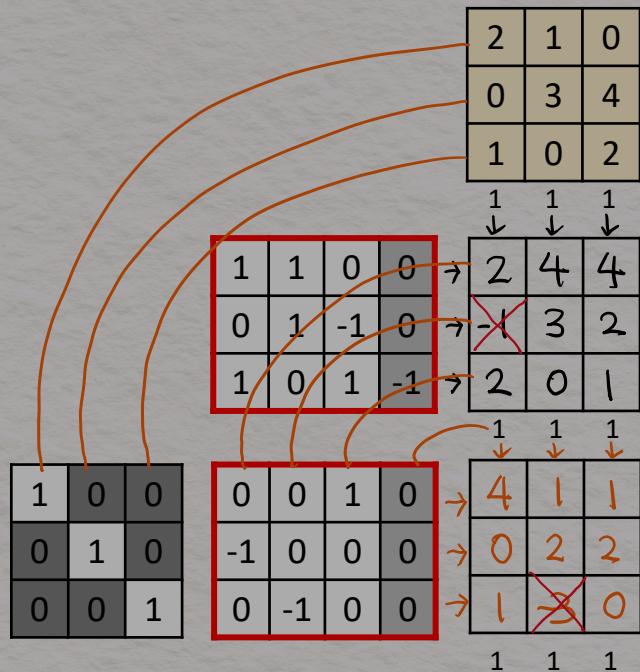
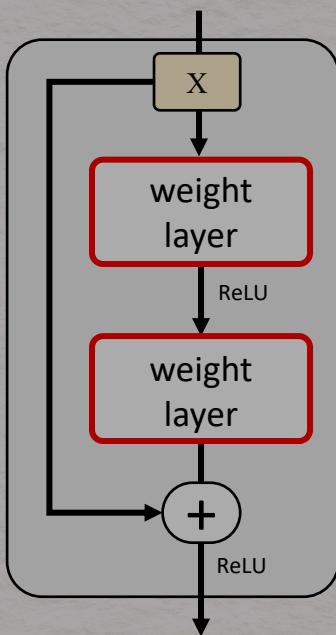
Residual Network



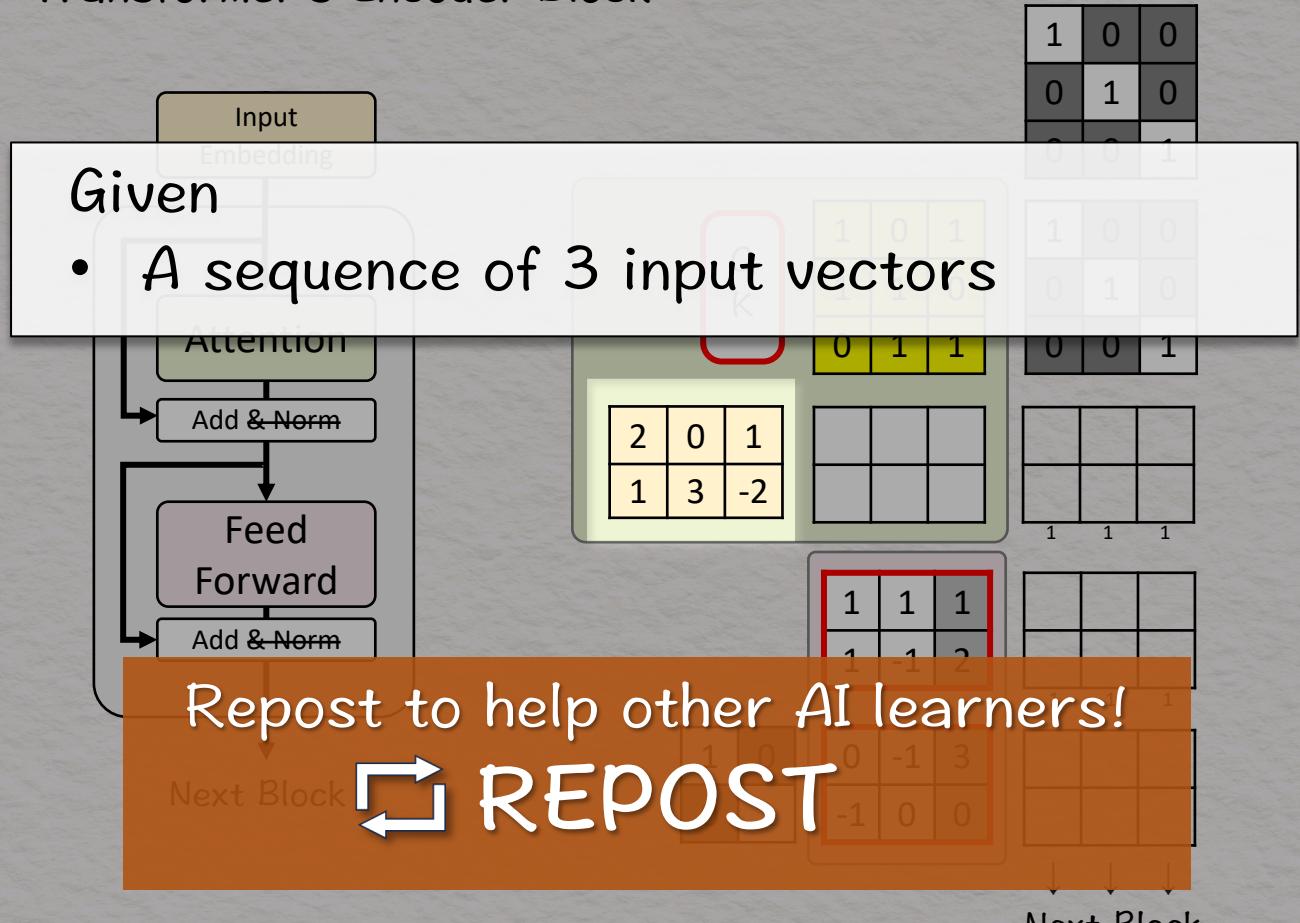
Transformer's Encoder Block



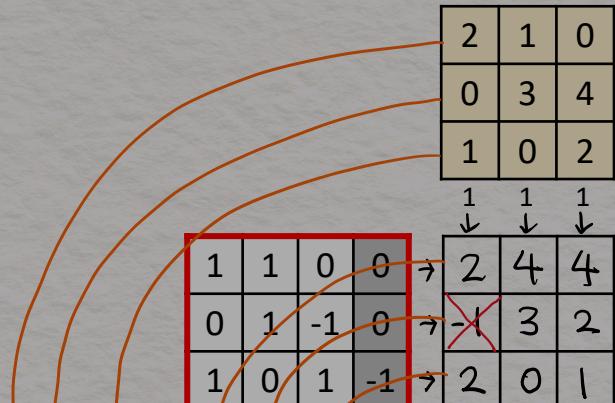
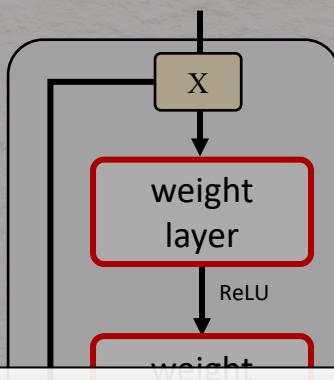
Residual Network



Transformer's Encoder Block

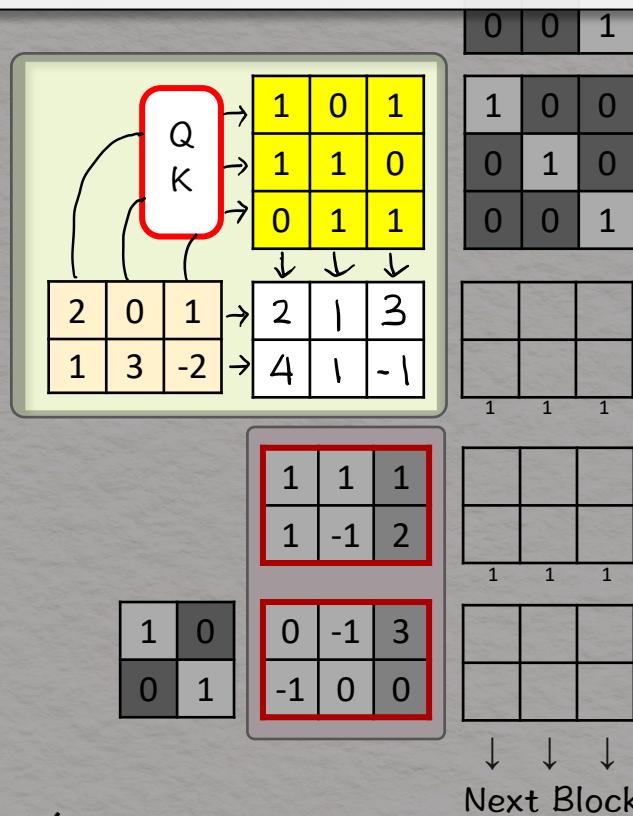
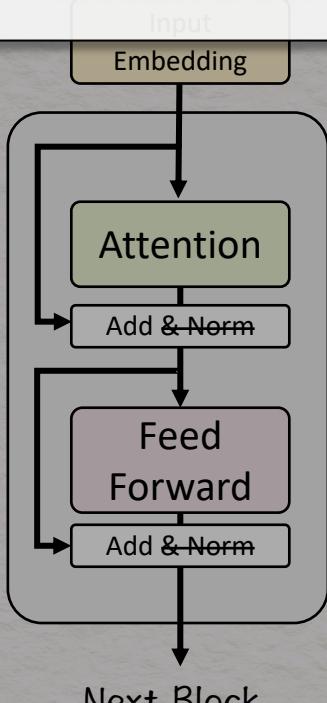


Residual Network

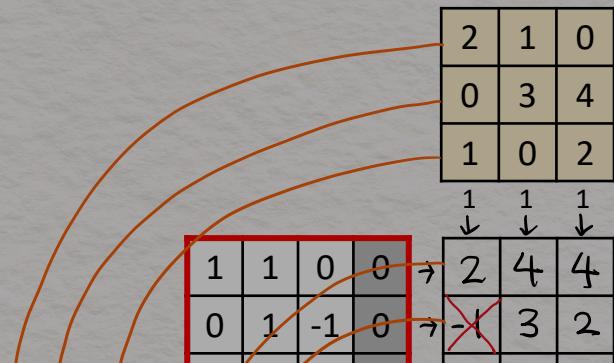
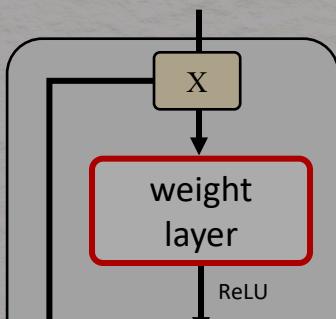


Attention

- Use the Query-Key module to compute a 3×3 attention matrix
- Multiply the input vectors with the attention matrix to obtain attention-weighted vectors

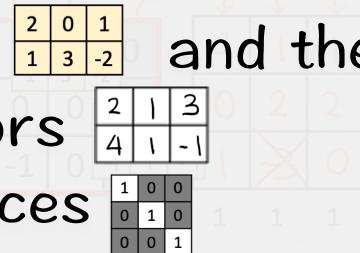


Residual Network

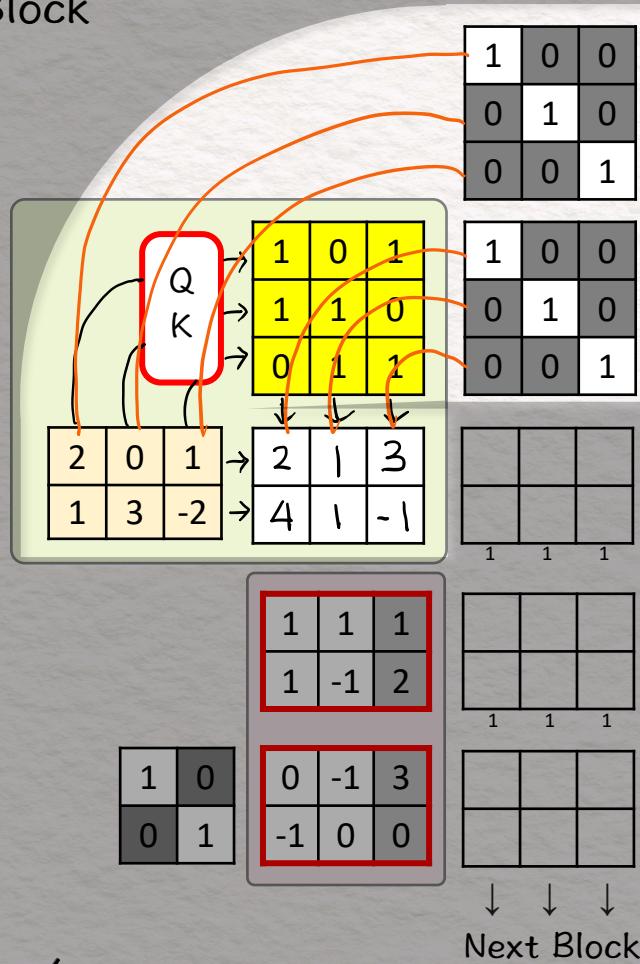
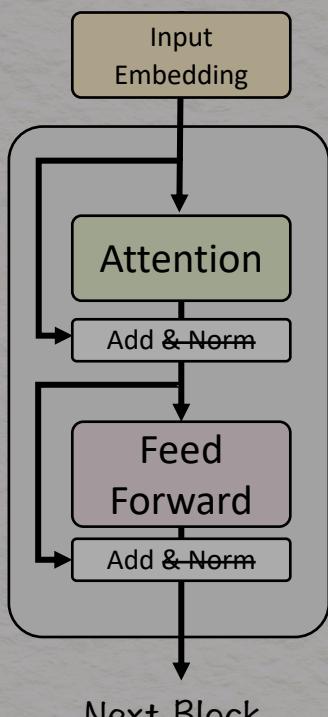


Concatenate

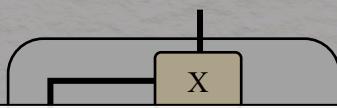
- Stack the input vectors and the attention-weighted vectors
- Stack two identity matrices



Transformer's Encoder Block



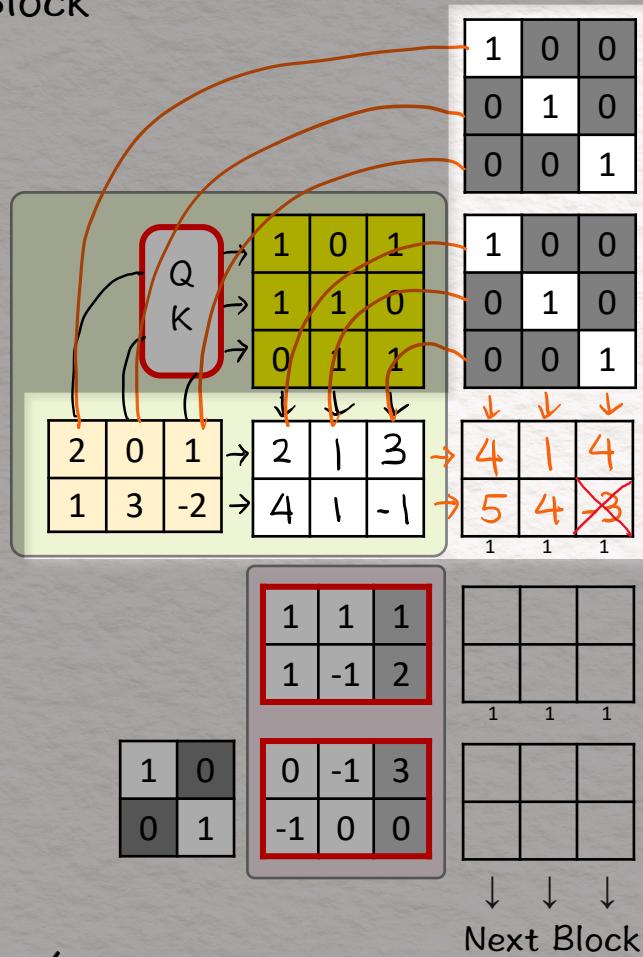
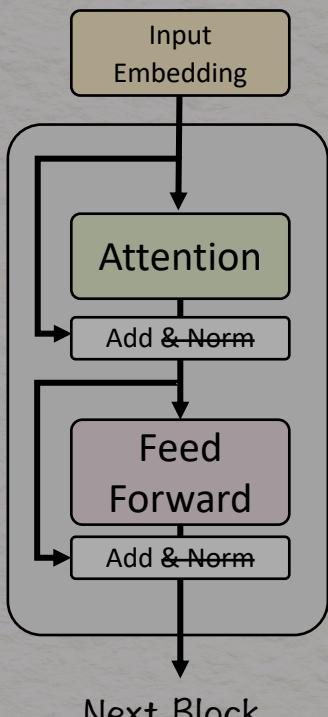
Residual Network



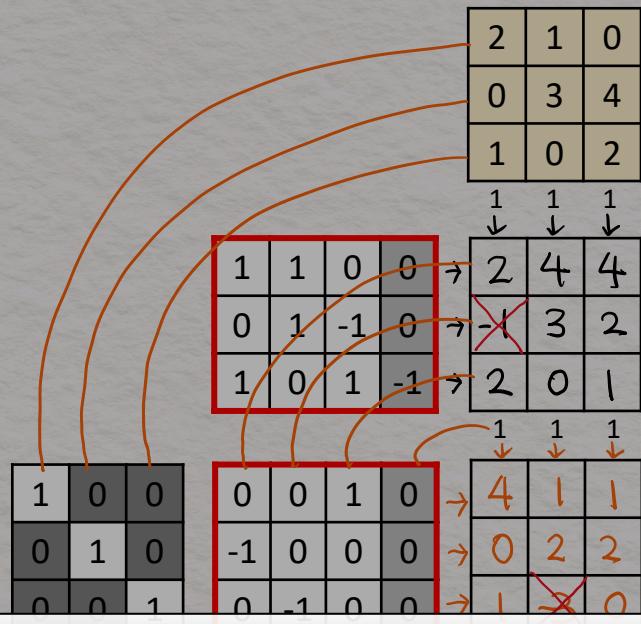
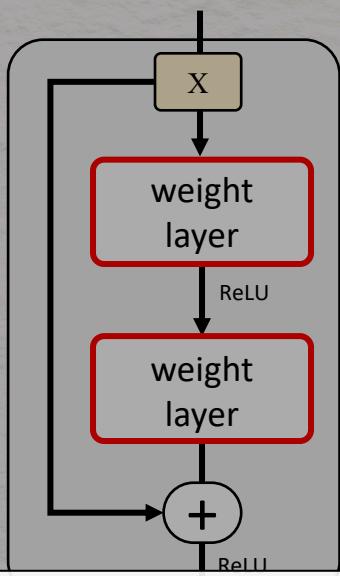
Add

- Multiply the two stacked matrices, which is equivalent to adding the input vectors $\begin{bmatrix} 2 & 0 & 1 \\ 1 & 3 & -2 \end{bmatrix}$ and the attention weighted vectors $\begin{bmatrix} 2 & 1 & 3 \\ 4 & 1 & -1 \end{bmatrix}$
- Apply ReLU (negatives $\rightarrow 0$)
- Obtain 3 new feature vectors $\begin{bmatrix} 4 & 1 & 4 \\ 5 & 4 & \cancel{3} \end{bmatrix}$

Transformer's Encoder Block



Residual Network



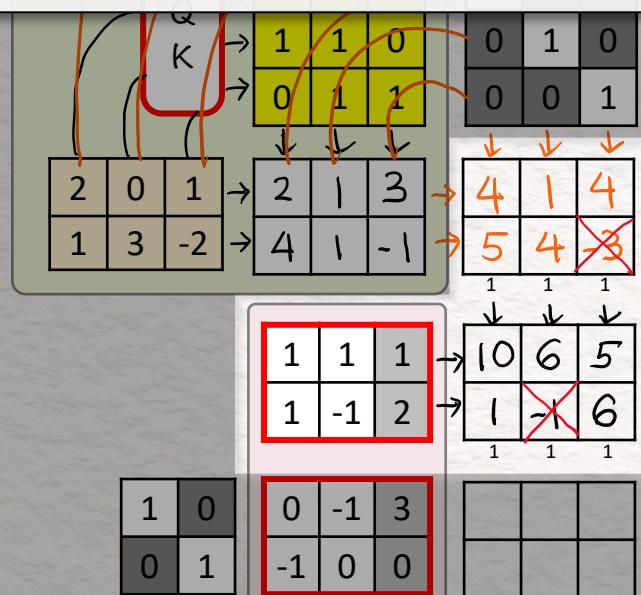
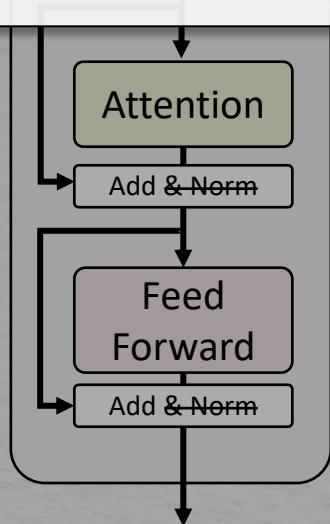
Feed Forward: First Layer

- Multiply the 3 feature vectors with weights and biases
- Apply ReLU (negatives \rightarrow 0)
- Obtain 3 new feature vectors

| | | |
|---|---|--------------|
| 4 | 1 | 4 |
| 5 | 4 | 3 |

| | | |
|---|----|---|
| 1 | 1 | 1 |
| 1 | -1 | 2 |

| | | |
|----|--------------|---|
| 10 | 6 | 5 |
| 1 | 4 | 6 |

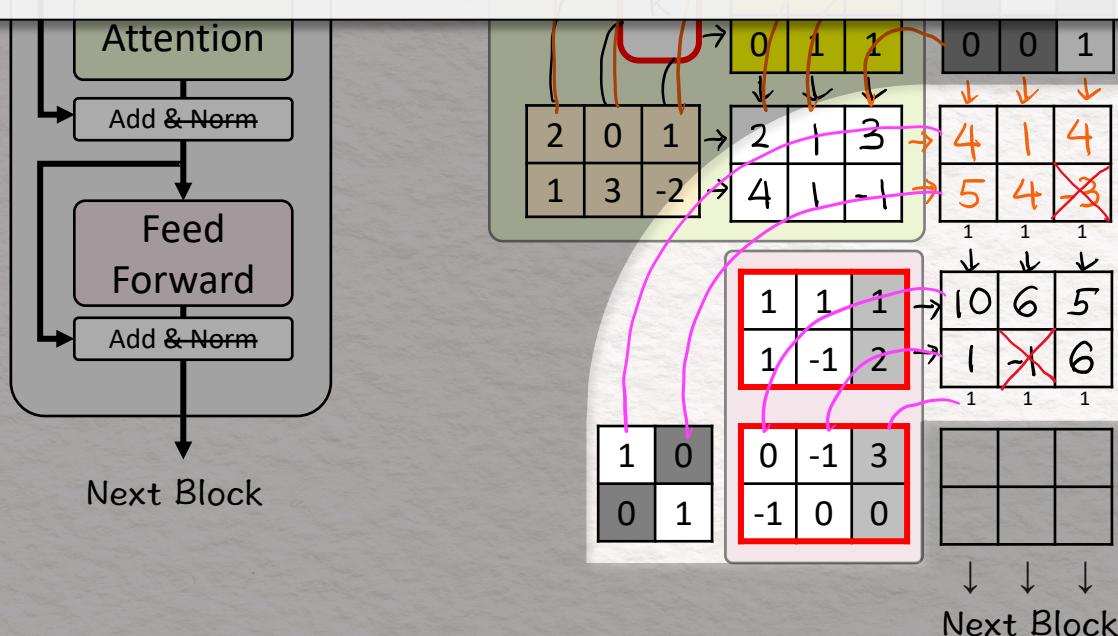


Residual Network

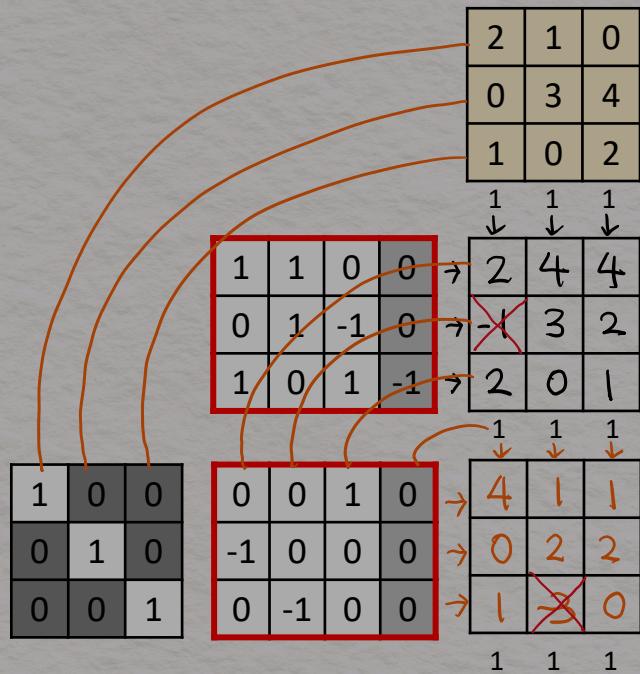
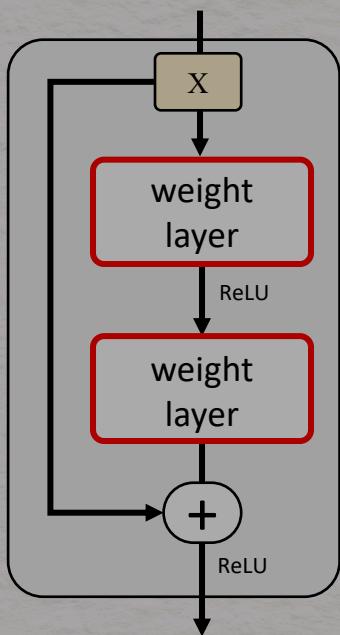


Feed Forward: Concatenate

- Stack the feature vectors from the attention layer $\begin{bmatrix} 4 & 1 & 4 \\ 5 & 4 & \cancel{4} \end{bmatrix}$ and the feature vectors from the first feed forward layer $\begin{bmatrix} 10 & 6 & 5 \\ 1 & \cancel{4} & 6 \end{bmatrix}$
- Stack an identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and the weight and bias matrix $\begin{bmatrix} 0 & -1 & 3 \\ -1 & 0 & 0 \end{bmatrix}$ of the second layer

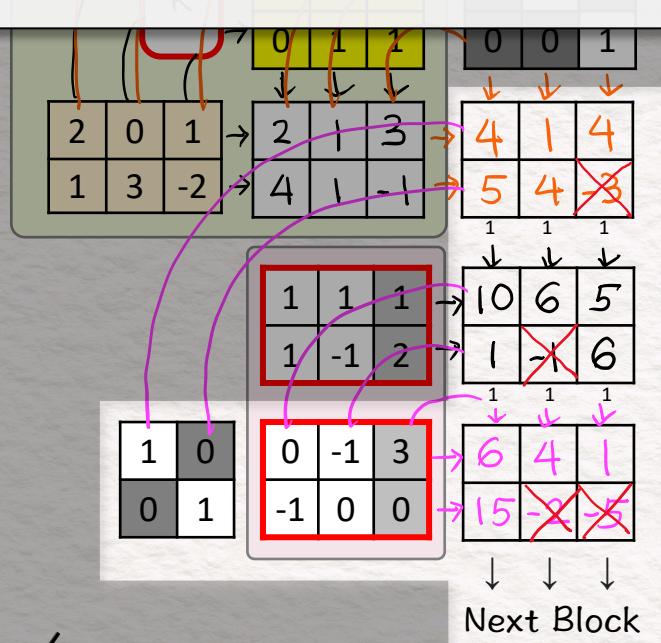
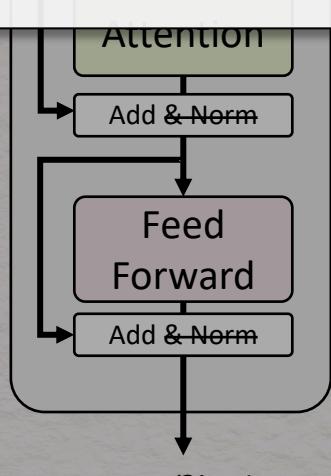


Residual Network



Feed Forward: Second Layer + Identity

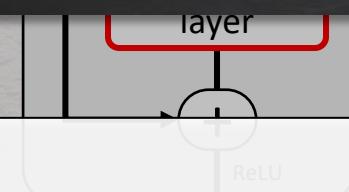
- Multiply the two stacked matrices
- Apply ReLU (negatives $\rightarrow 0$)
- Obtain new feature vectors
- Pass them to the next encoder block



Residual Network

| | | |
|---|---|---|
| 2 | 1 | 0 |
|---|---|---|

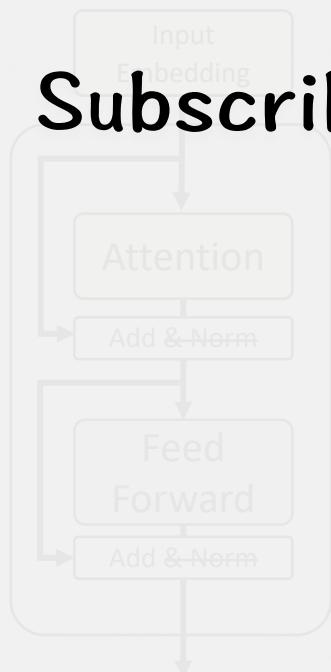
AI by Hand 🖋️ Advanced Series



| | | | | | | | | | | | |
|---|---|---|---|---|---|---|----|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

Never miss a future post in
the series

Transformer's Encoder Block



<http://by-hand.ai/news>