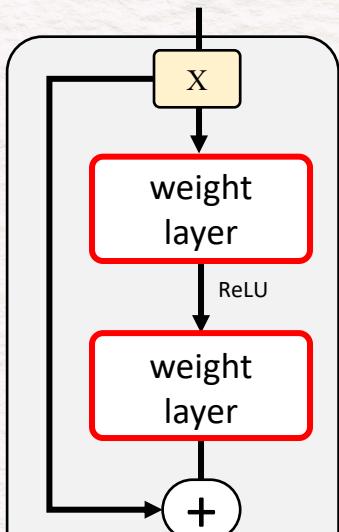
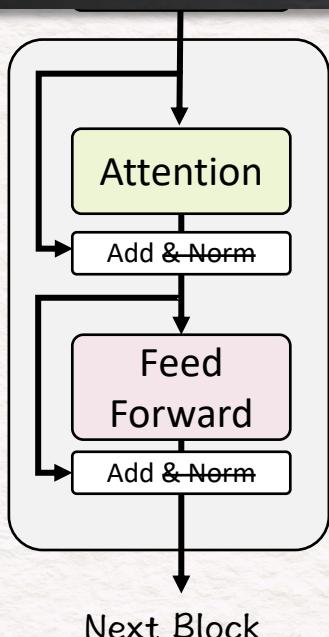


Residual Network



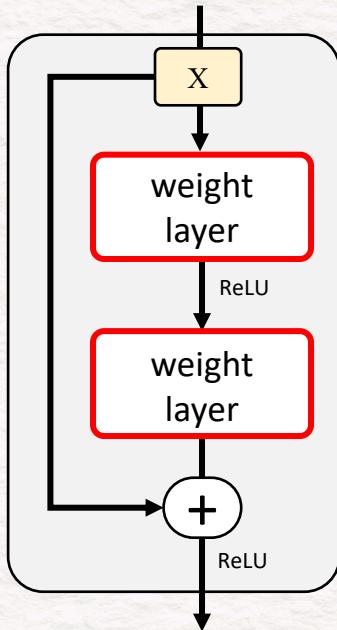
$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 4 \\ 1 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$
$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$
$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$

AI by Hand Advanced Series



$\begin{bmatrix} Q \\ K \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$
$\begin{bmatrix} 2 & 0 & 1 \\ 1 & 3 & -2 \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$
$\begin{bmatrix} 0 & -1 & 3 \\ -1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$	$\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}$
	\downarrow	\downarrow
		\downarrow
		Next Block

Residual Network



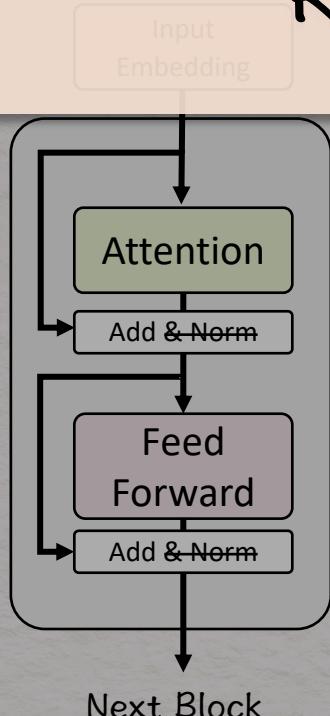
	2	1	0
	0	3	4
	1	0	2
1	1	1	1

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

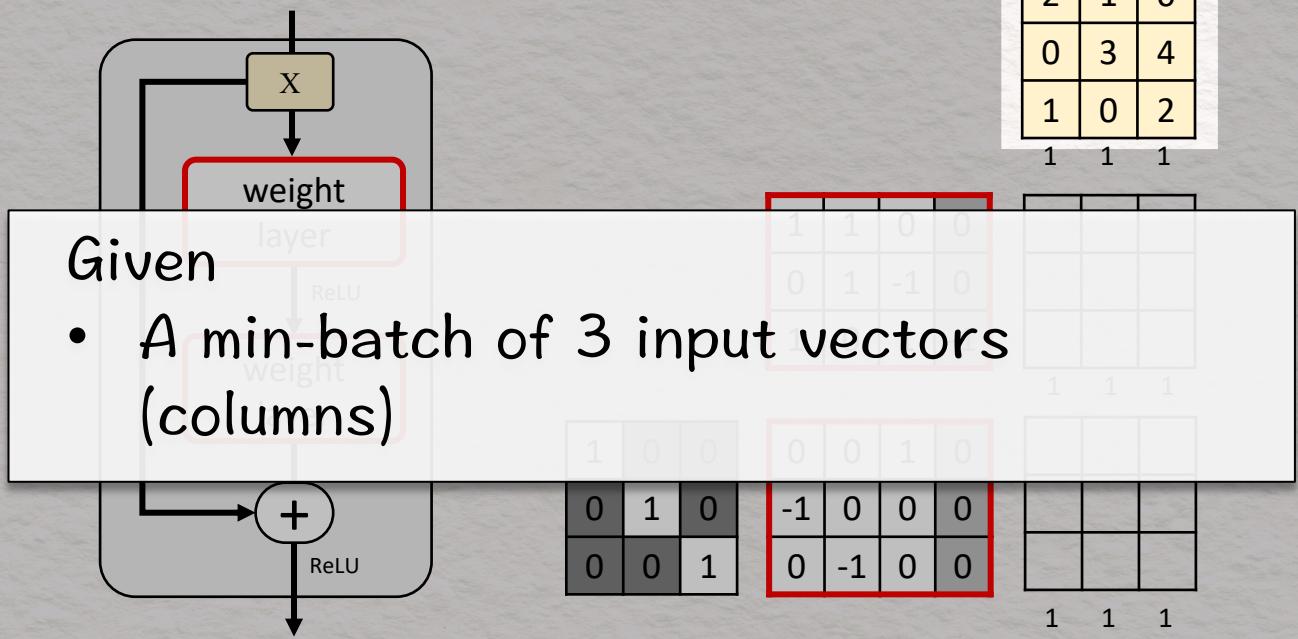
Residual Block



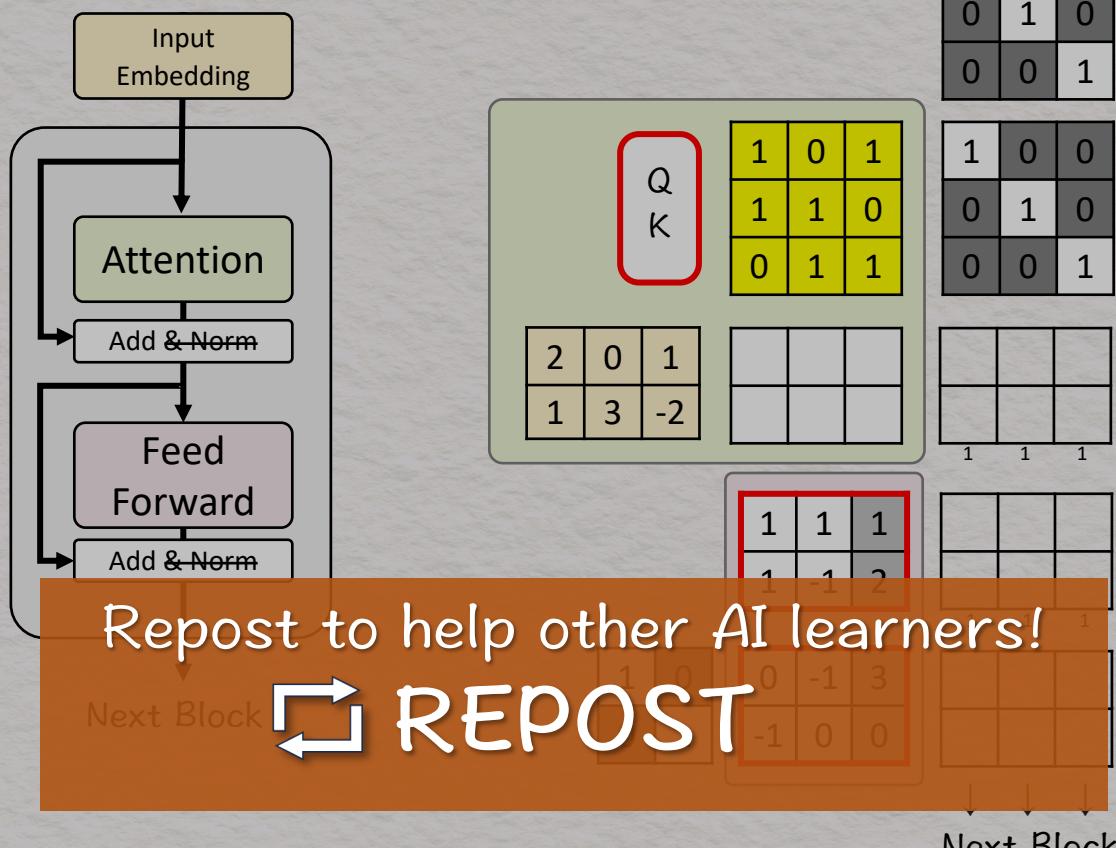
Q	1	0	1
K	1	1	0
	0	1	1
2	1	0	1
1	3	-2	
1	1	1	
1	-1	2	
1	0		
0	-1	3	
-1	0	0	
1	1	1	
1	1	1	
1	1	1	

↓ ↓ ↓
Next Block

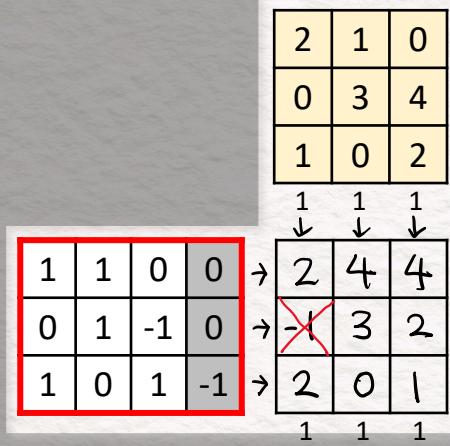
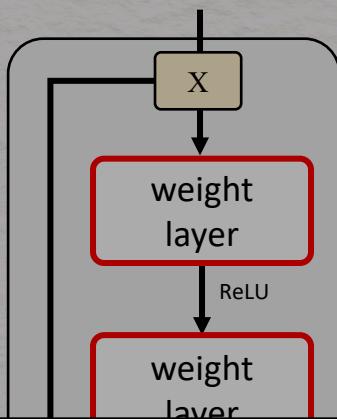
Residual Network



Transformer's Encoder Block

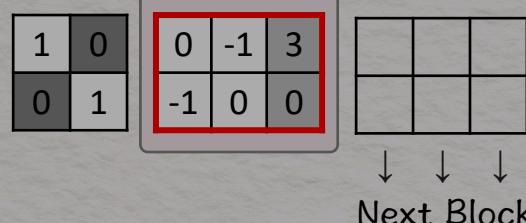
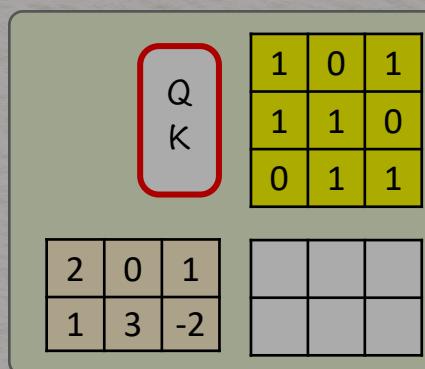
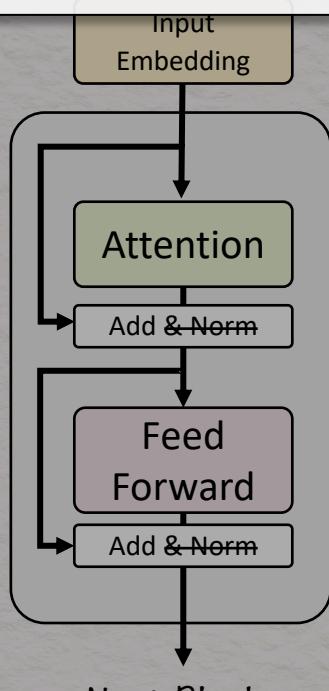


Residual Network

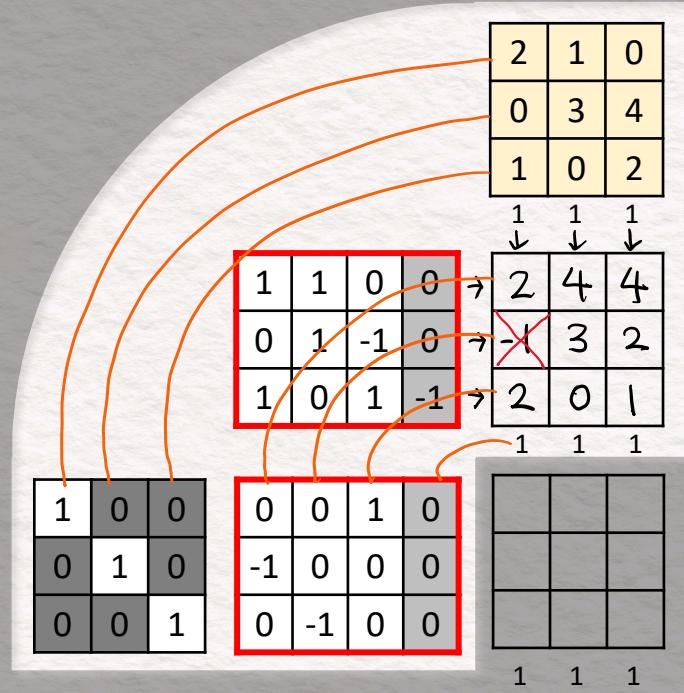
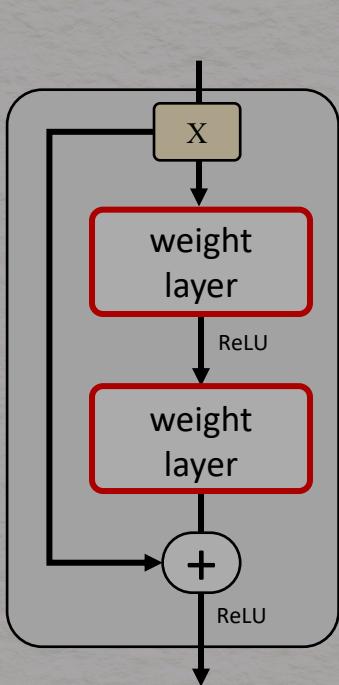


Linear Layer

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

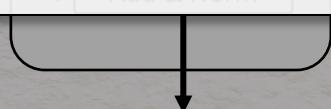


Residual Network

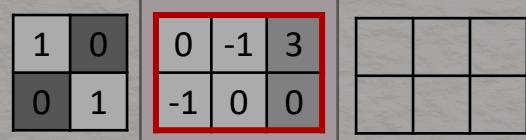


Concatenate

- 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)

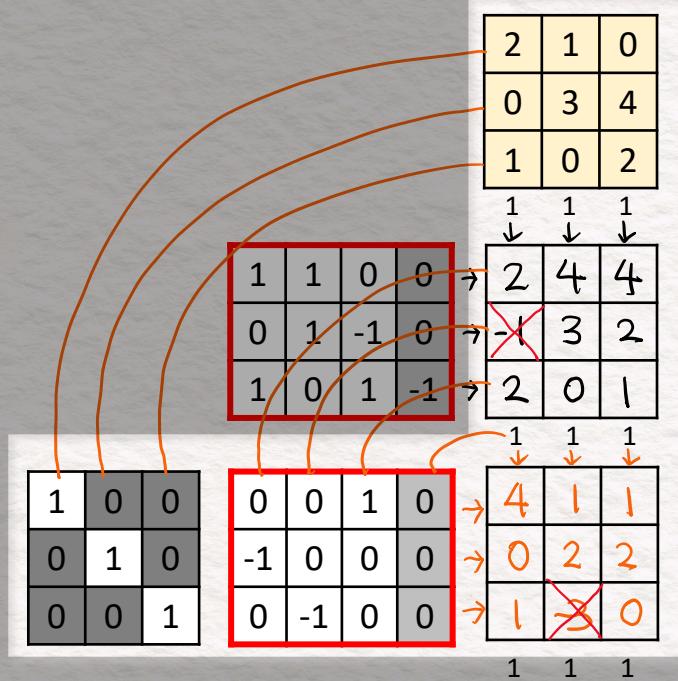
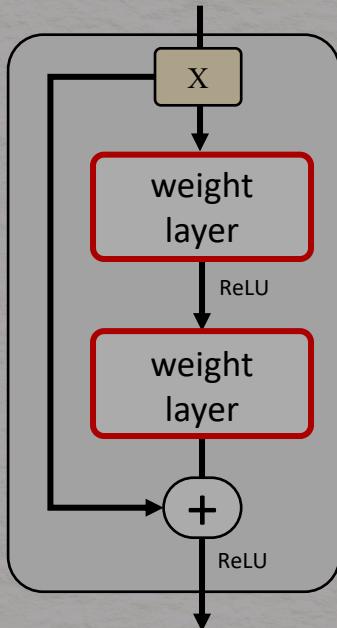


Next Block



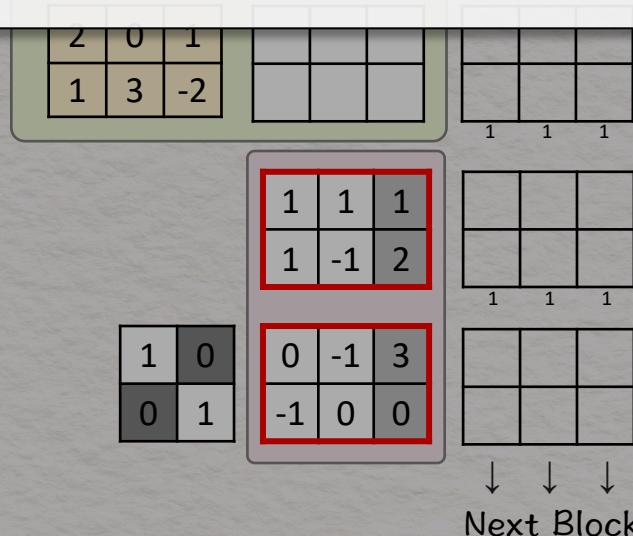
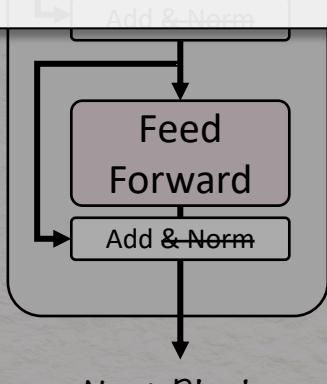
↓ ↓ ↓
Next Block

Residual Network

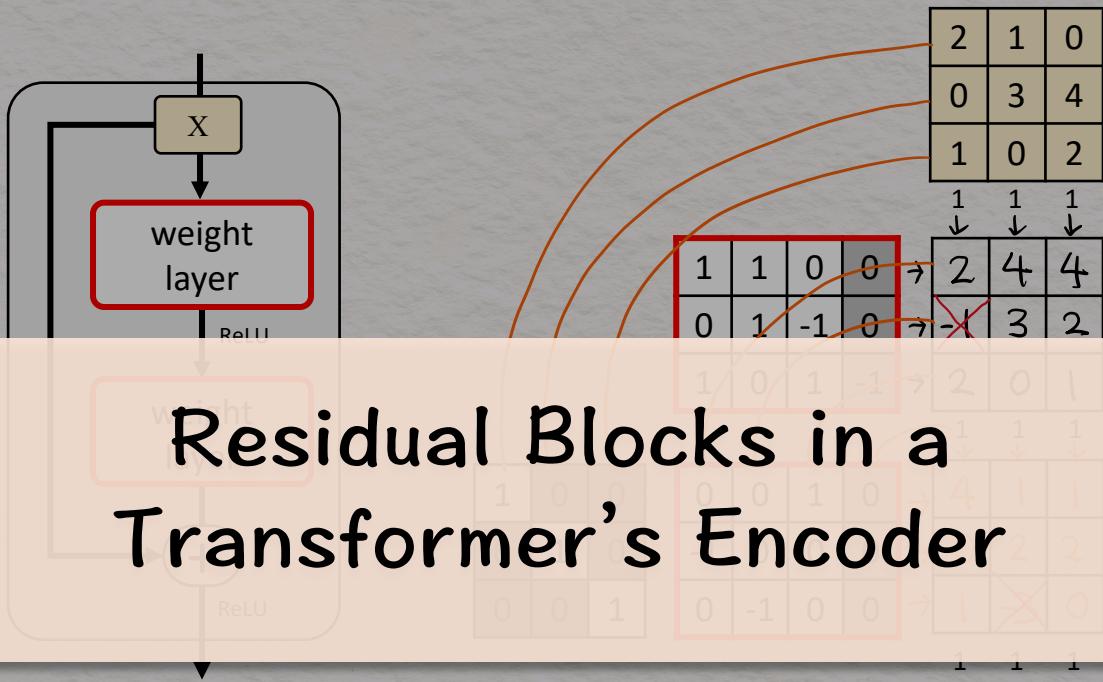


Linear Layer + Identity

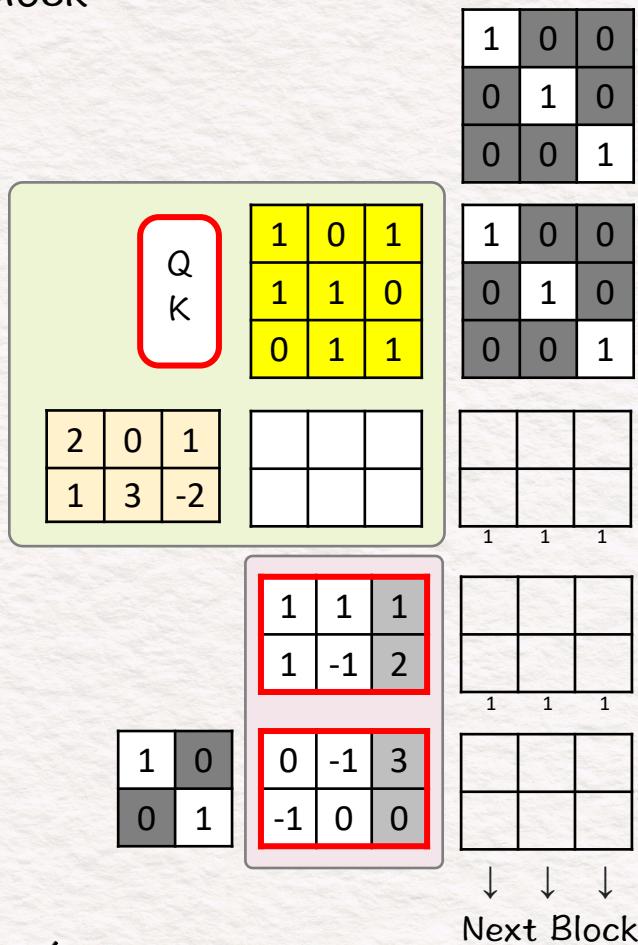
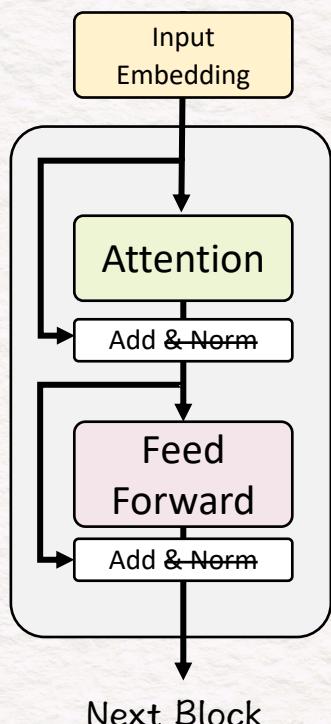
- Multiply the two stacked matrices
- This is equivalent to $F(X) + X$
- Apply ReLU (negatives $\rightarrow 0$)
- Obtain 3 new feature vectors
- Pass them to the next residual block



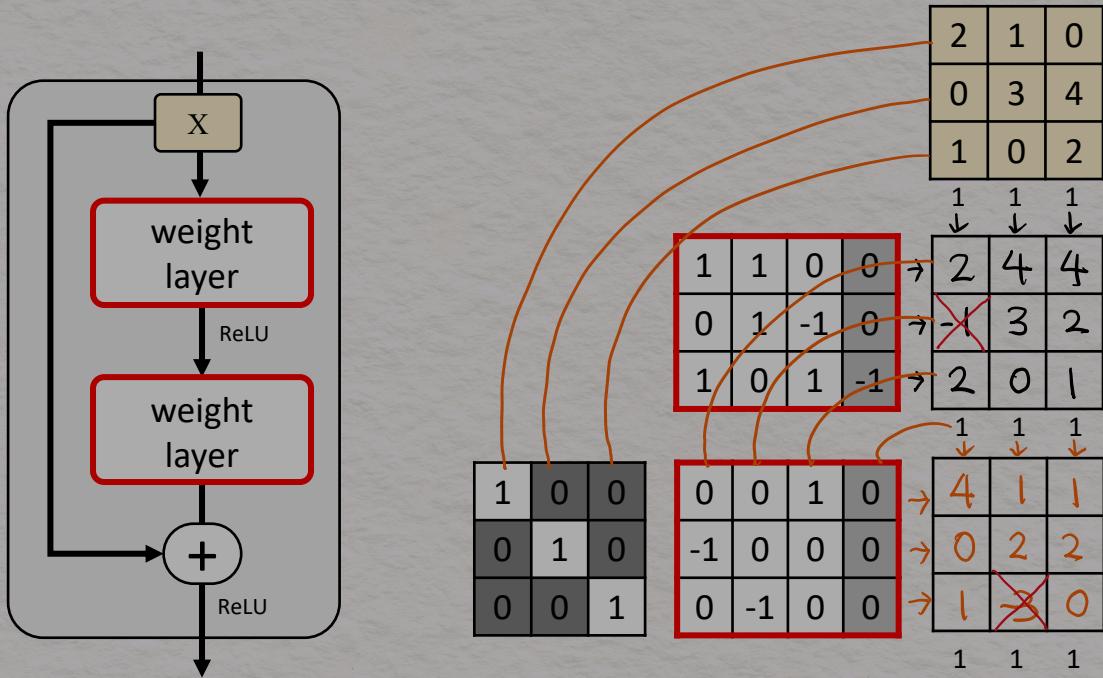
Residual Network



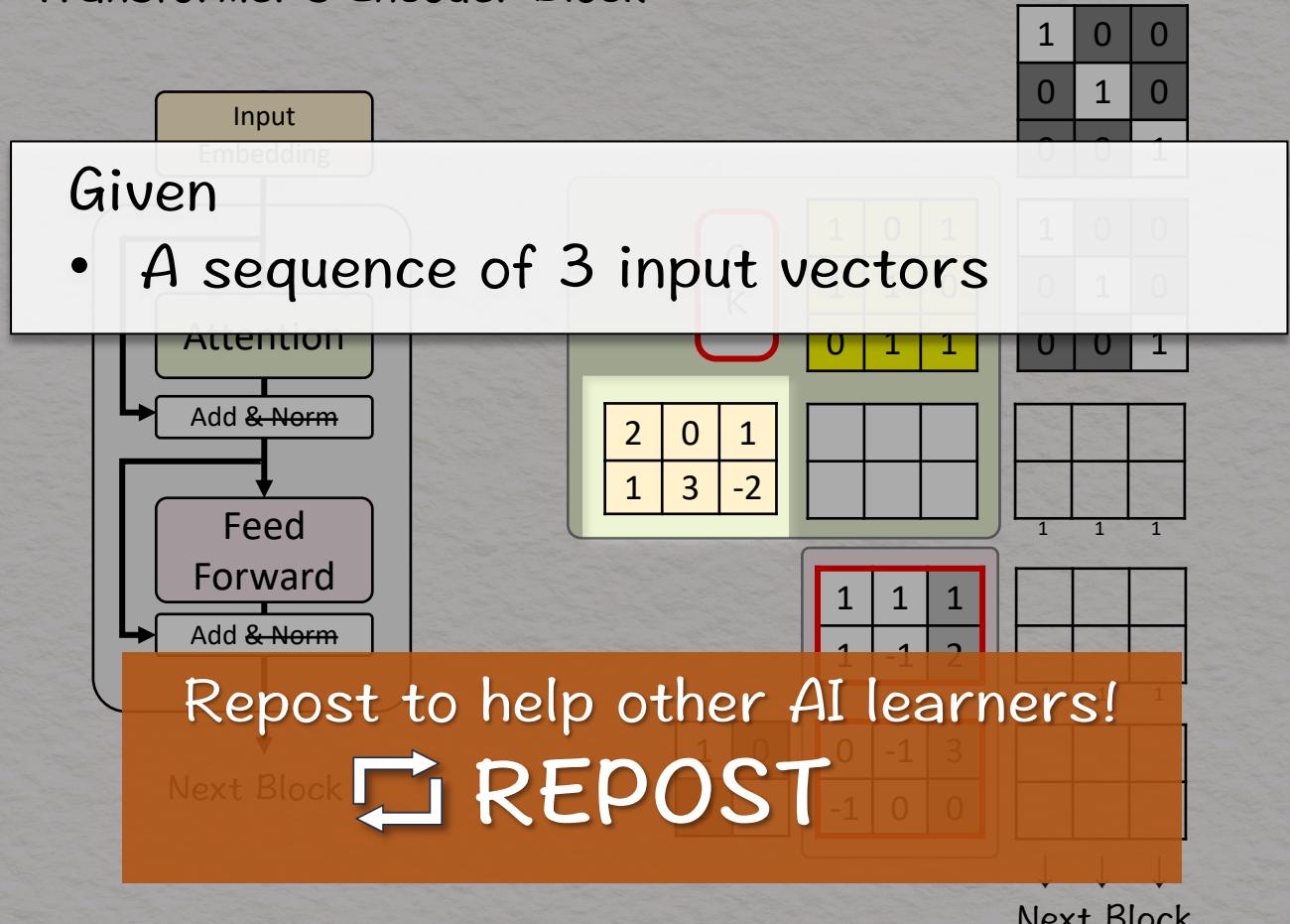
Transformer's Encoder Block



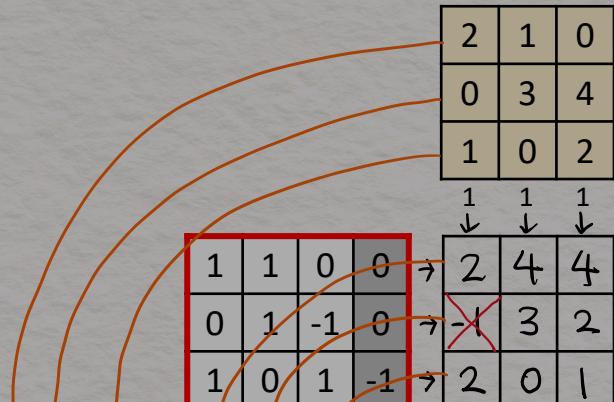
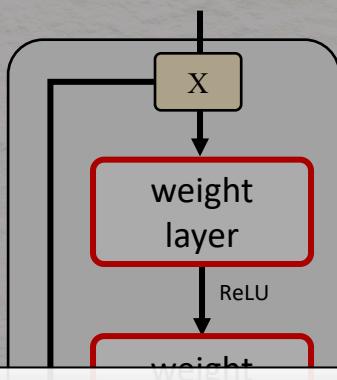
Residual Network



Transformer's Encoder Block

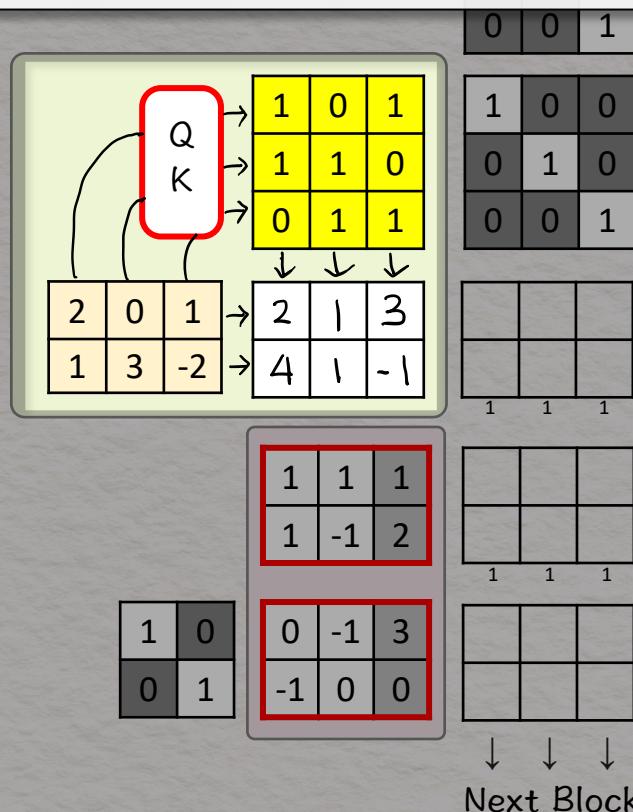
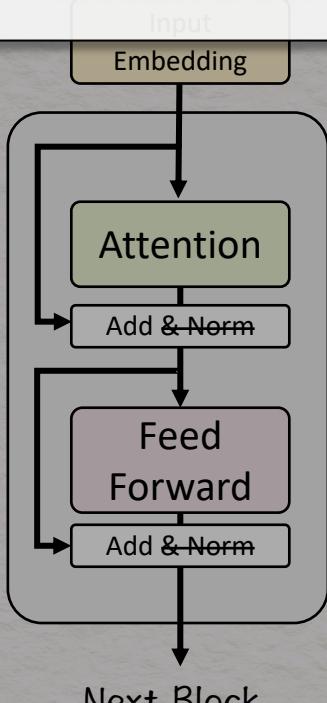


Residual Network

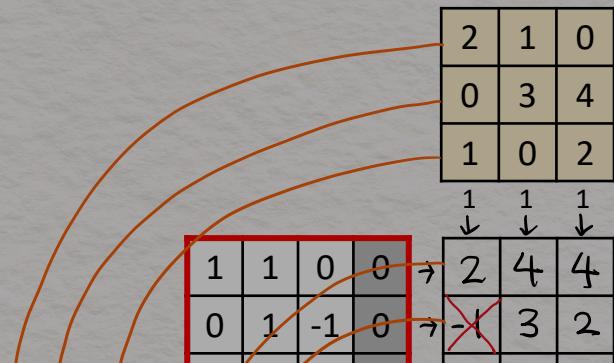
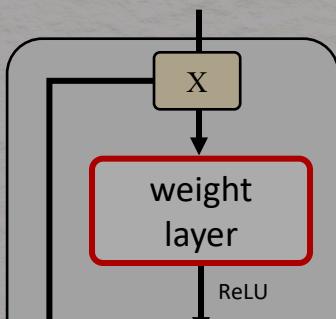


Attention

- 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

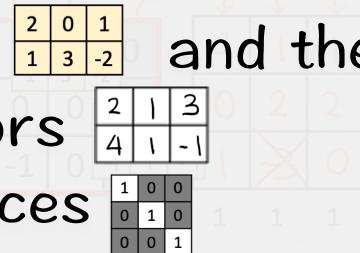


Residual Network

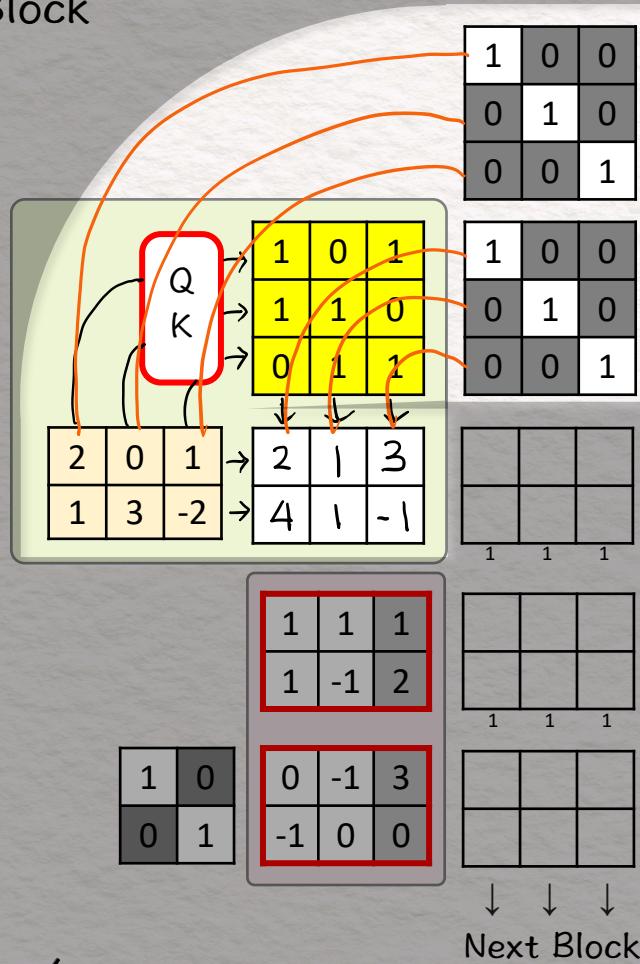
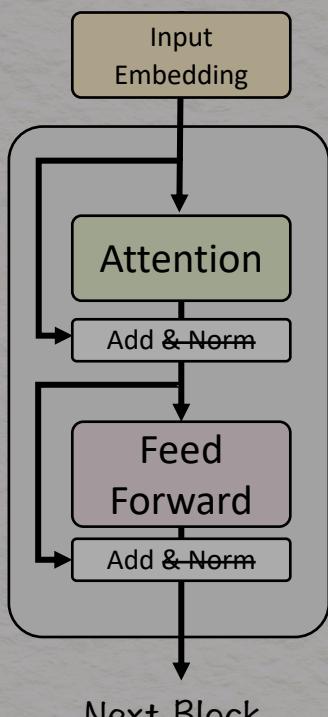


Concatenate

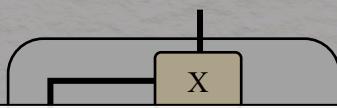
- Stack the input vectors and the attention-weighted vectors
- Stack two identity matrices



Transformer's Encoder Block



Residual Network

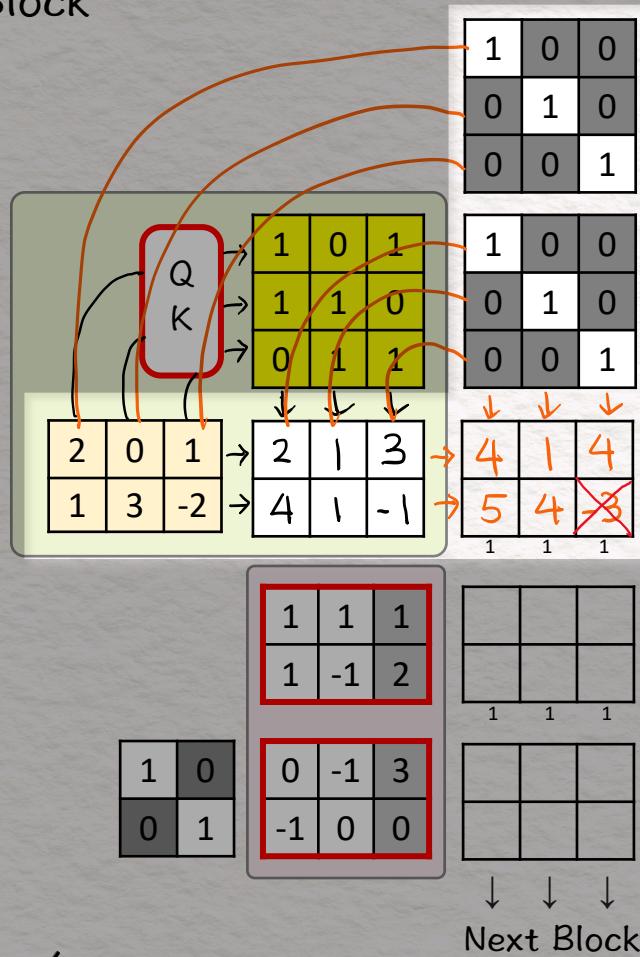
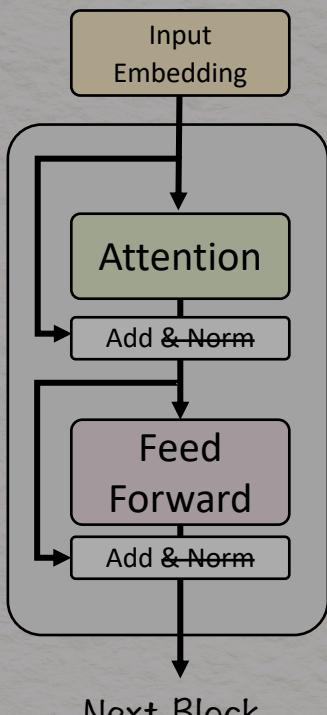


2	1	0
0	3	4
1	0	2

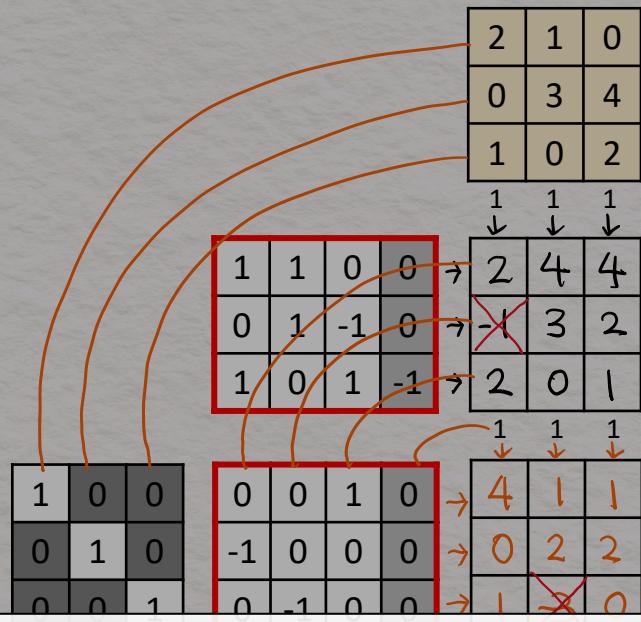
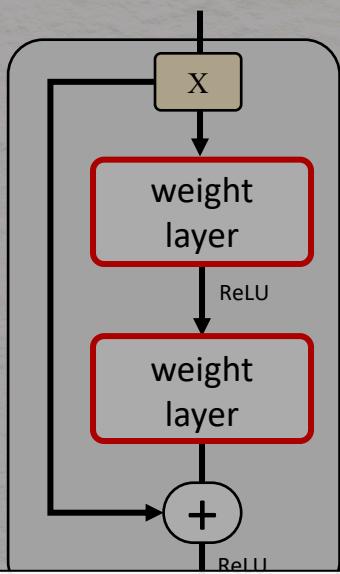
Add

- 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



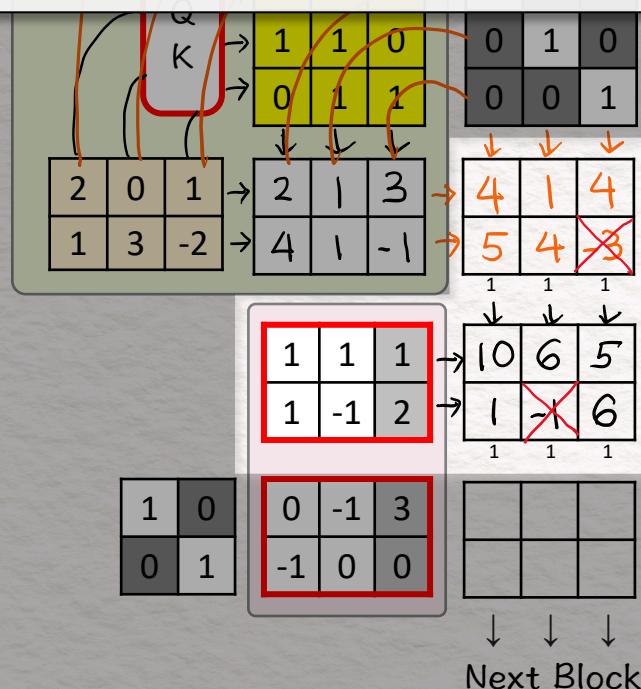
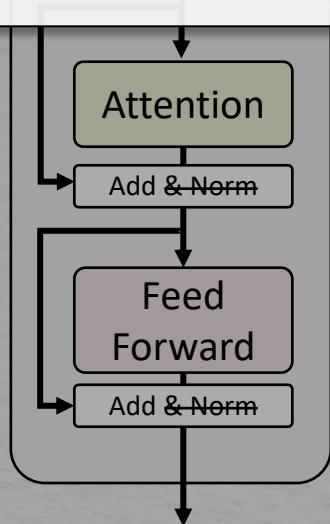
Feed Forward: First Layer

- 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

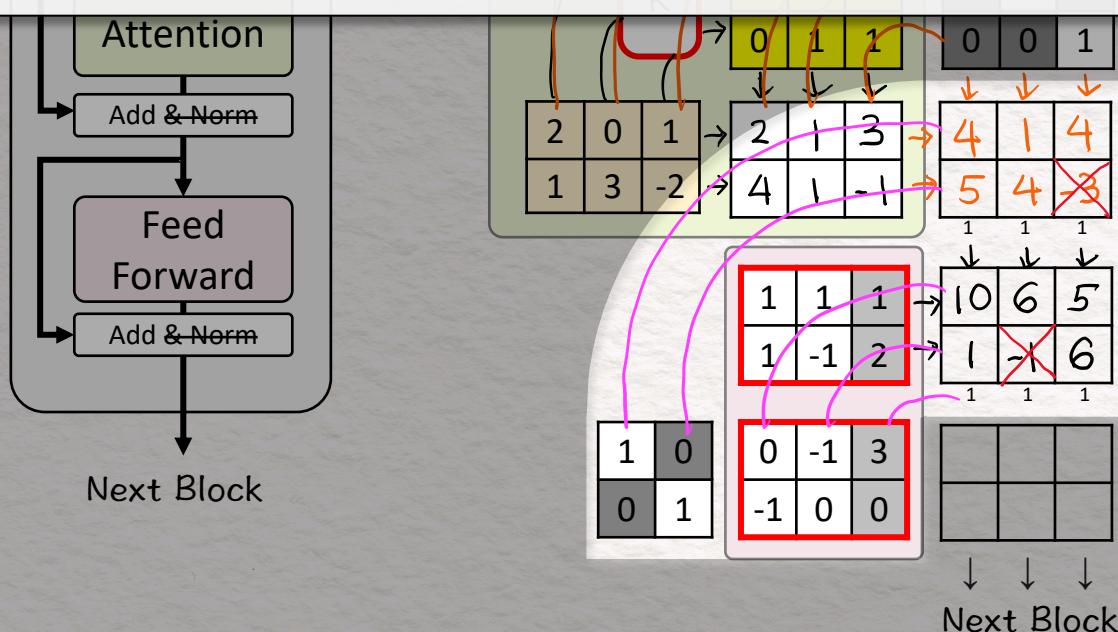


Residual Network

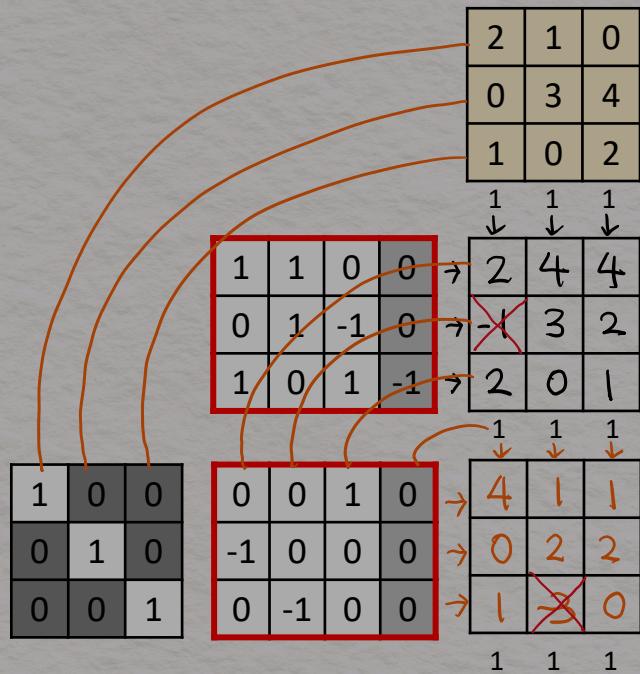
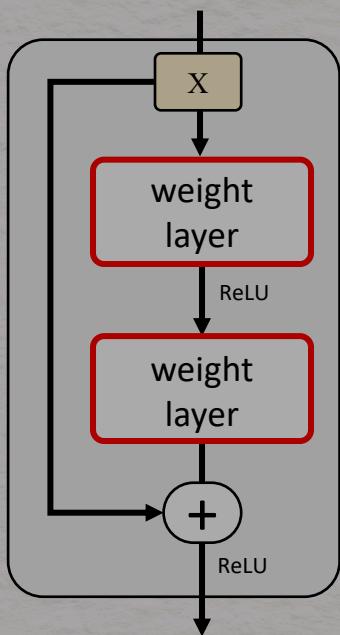


Feed Forward: Concatenate

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

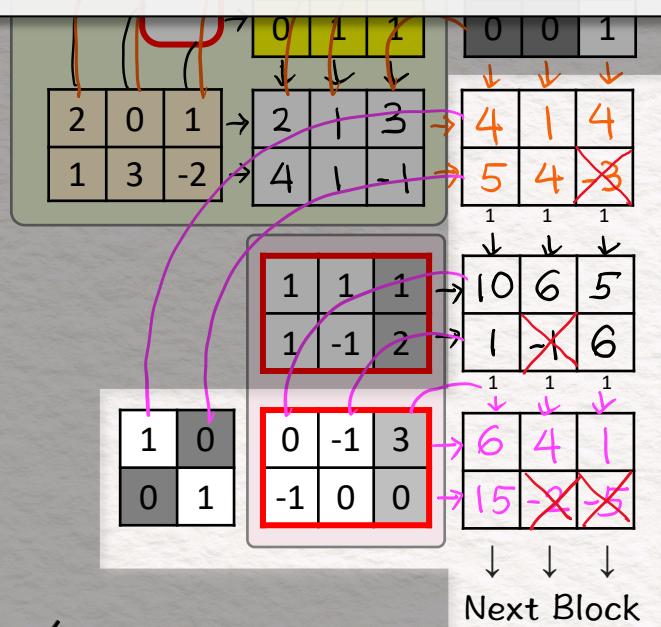
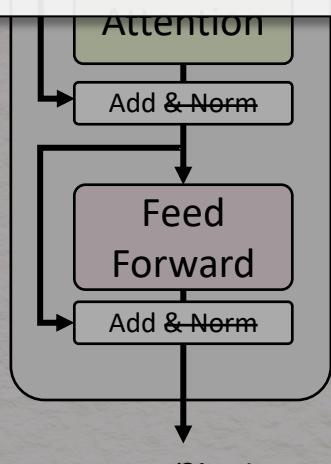


Residual Network



Feed Forward: Second Layer + Identity

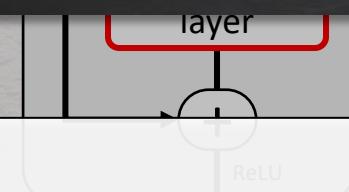
- Multiply the two stacked matrices
- Apply ReLU (negatives $\rightarrow 0$)
- Obtain new feature vectors
- Pass them to the next encoder block



Residual Network

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

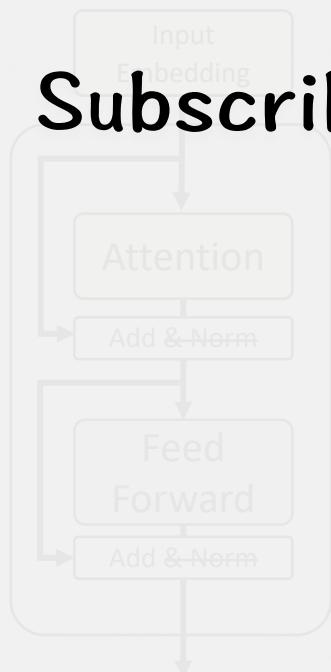
AI by Hand 🖋️ Advanced Series



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>