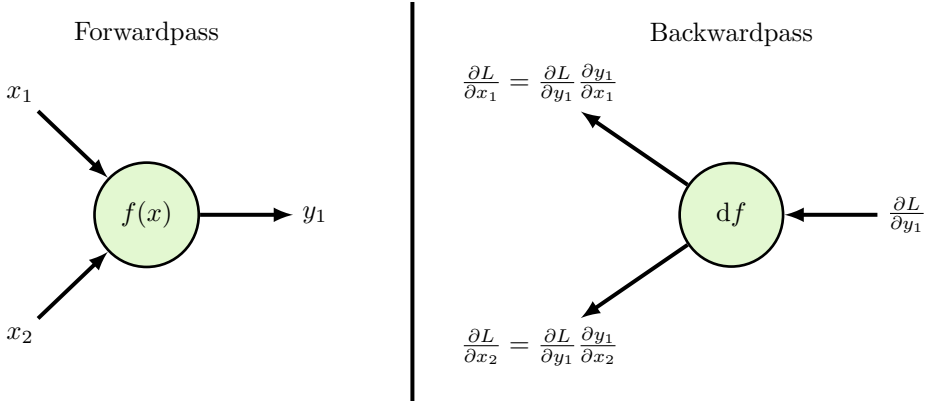
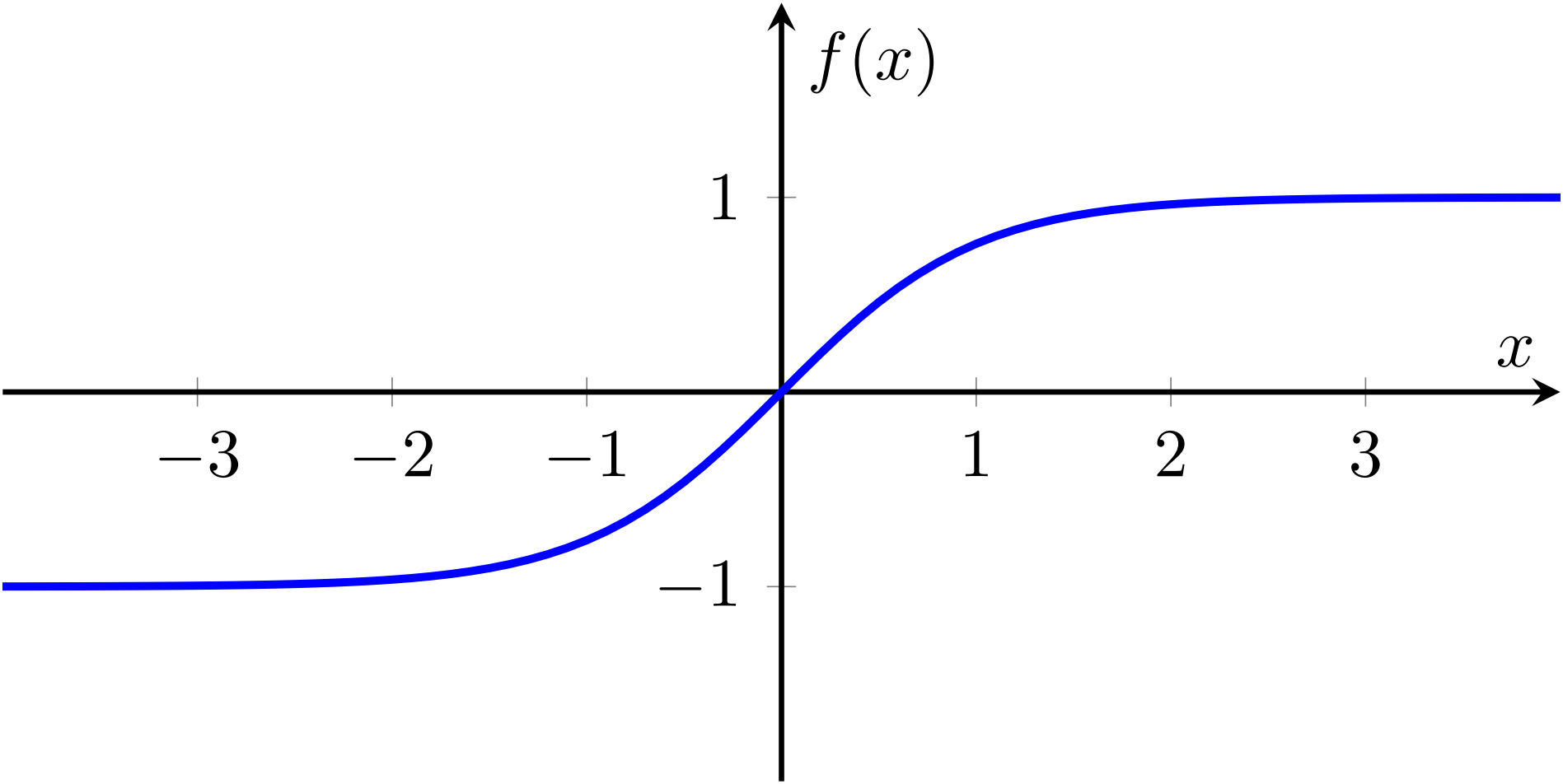
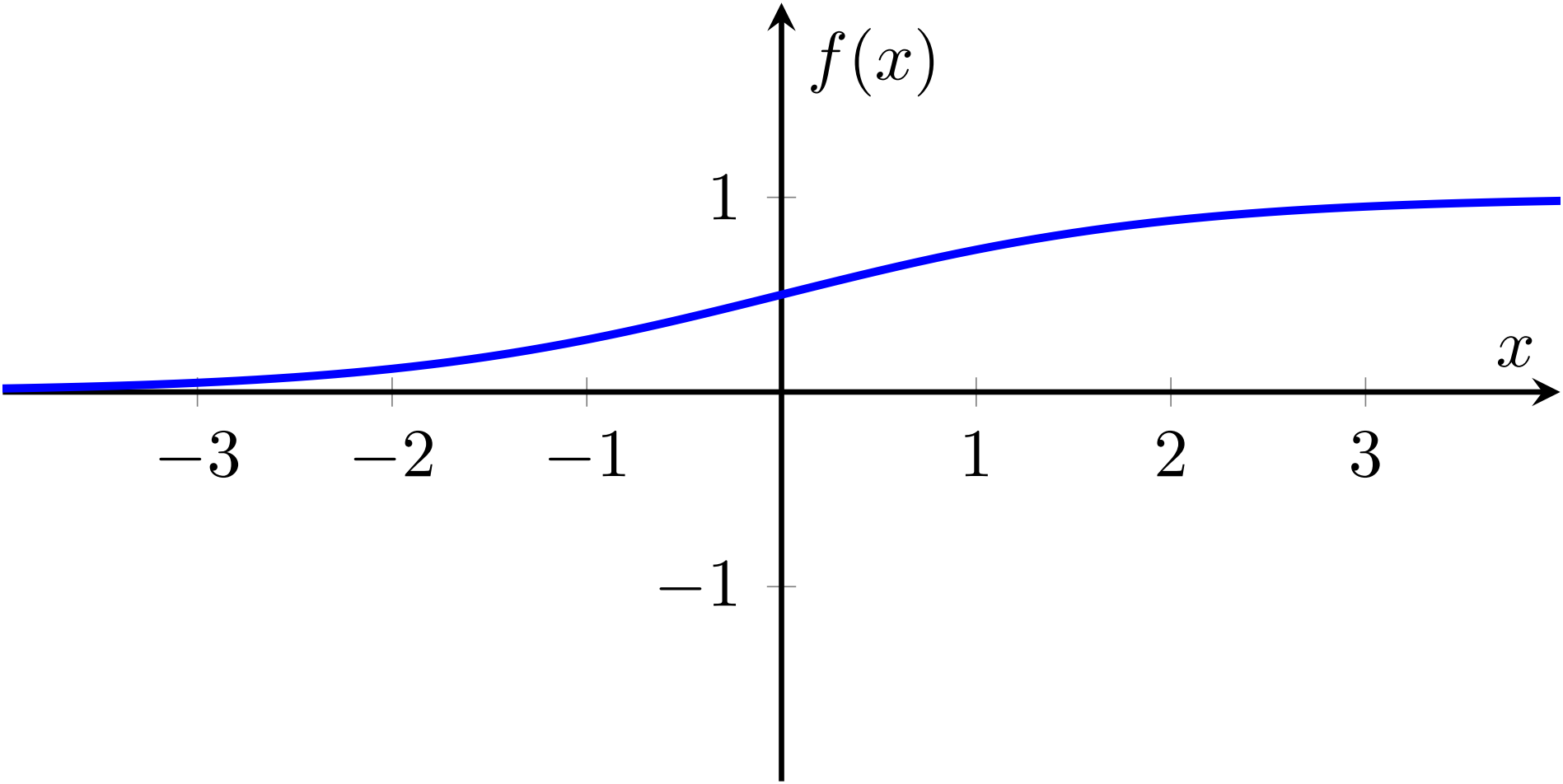
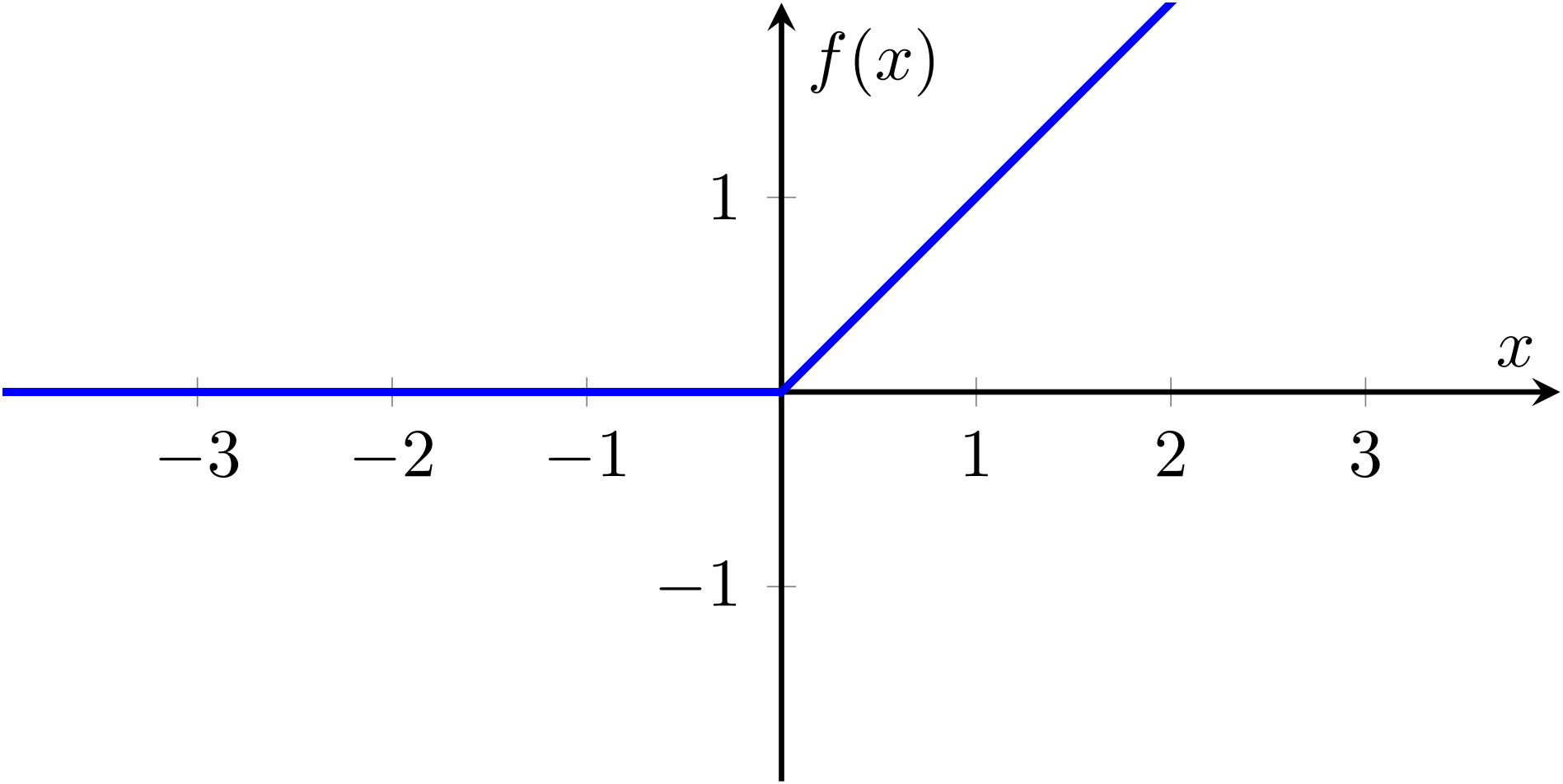
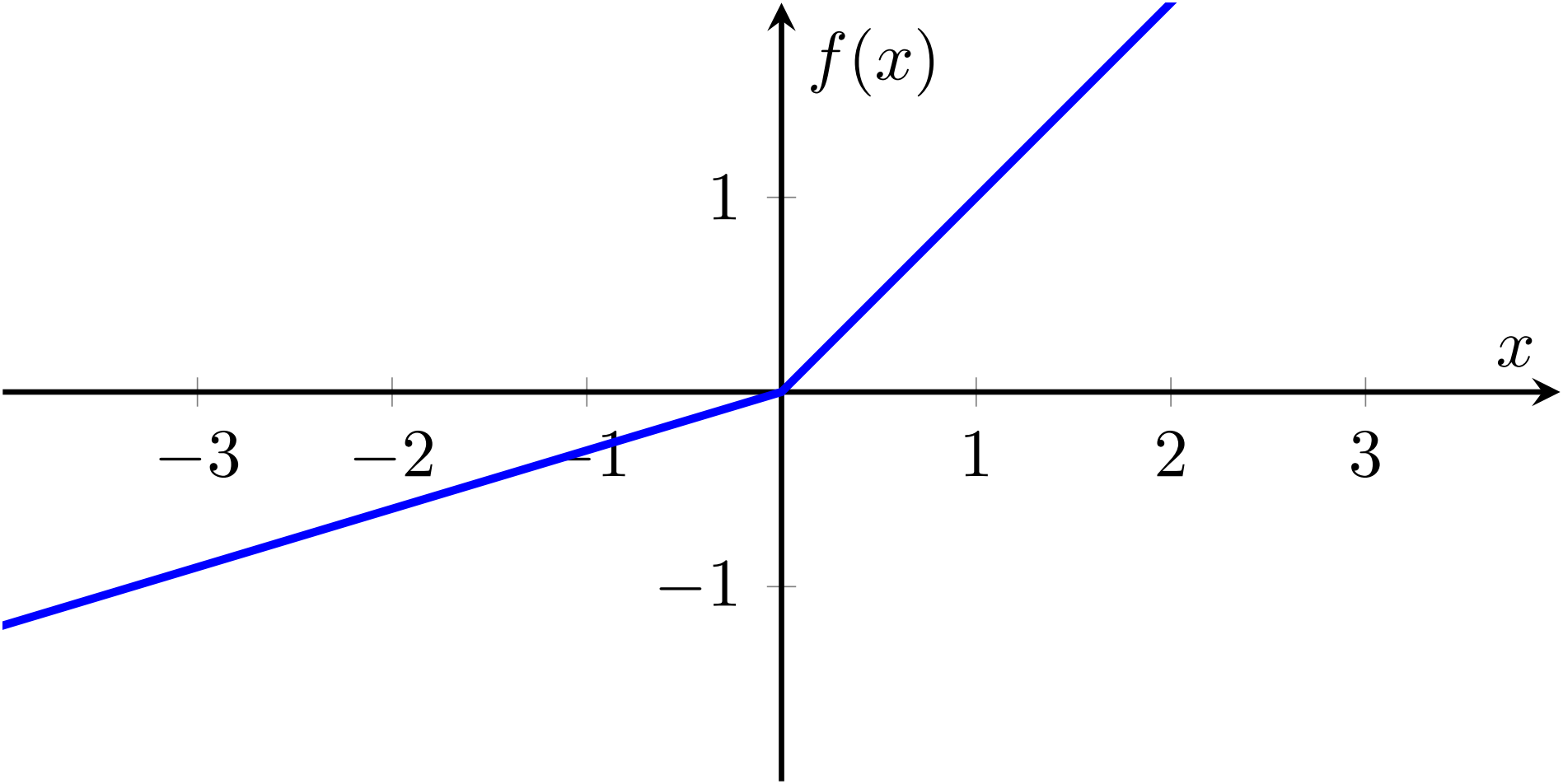
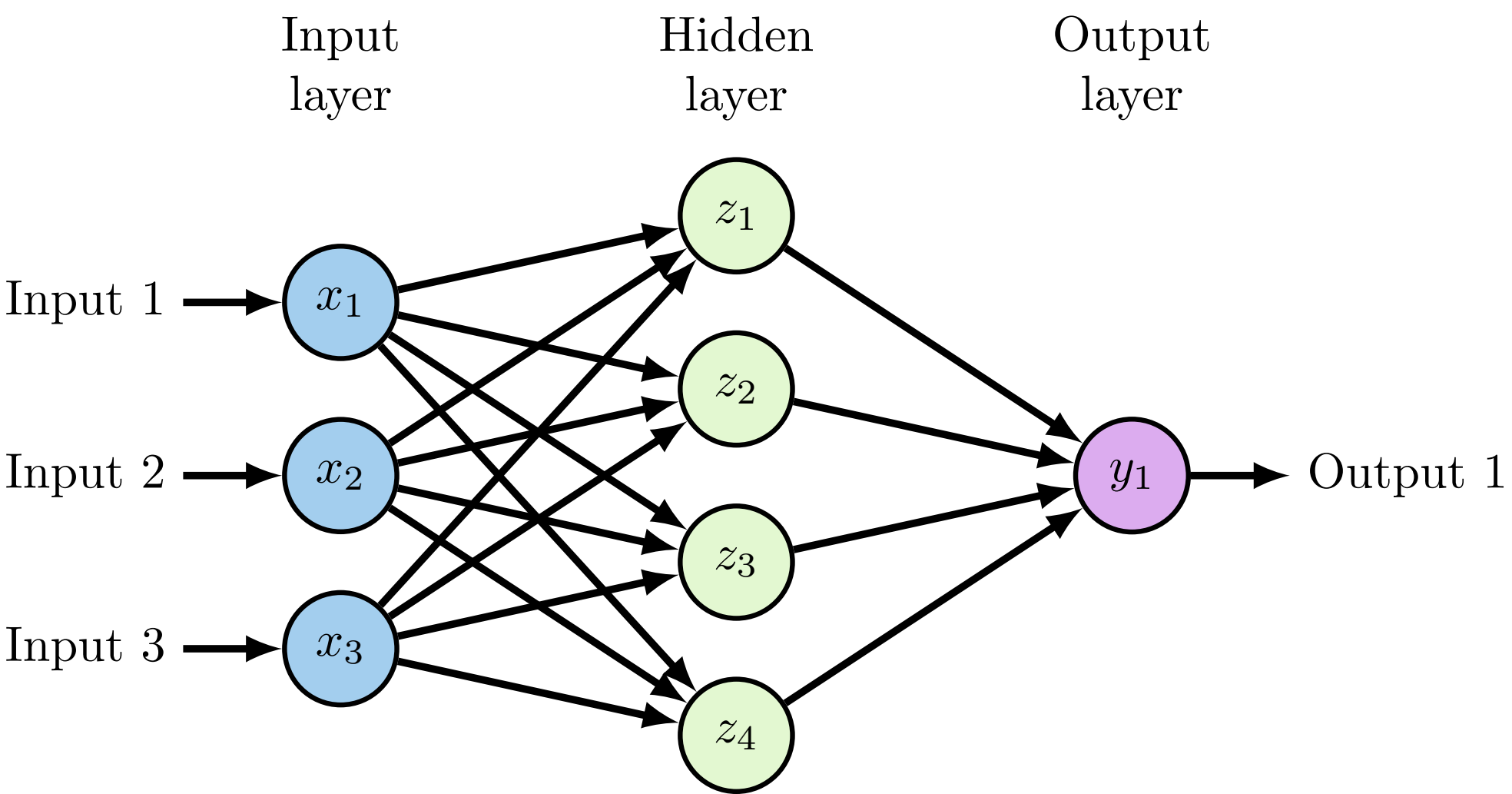

Forwardpass
x_1
x_2
$f(x)$
y_1
Backwardpass
$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial y_1}\frac{\partial y_1}{\partial x_1}$
$\mathrm{d}f$
$\frac{\partial L}{\partial y_1}$
$\frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial y_1}\frac{\partial y_1}{\partial x_2}$

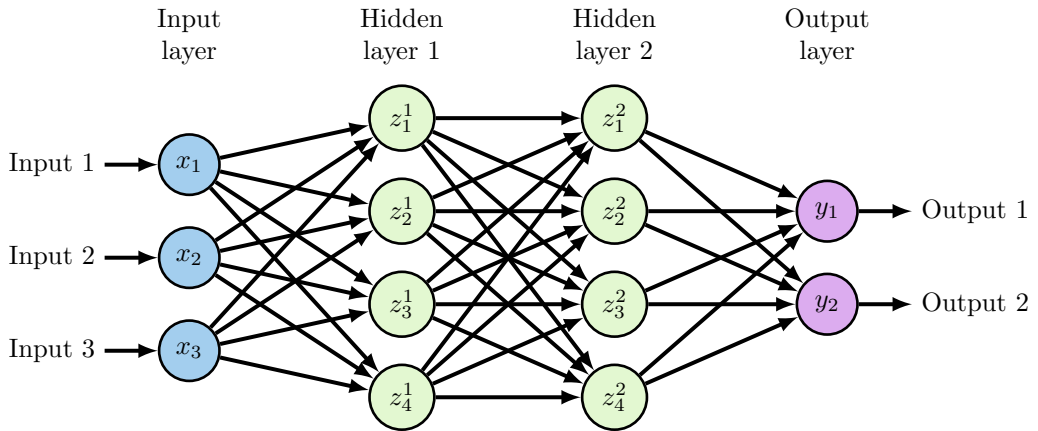

$f(x)$
x
1
-1
-3
-2
-1
1
2
3

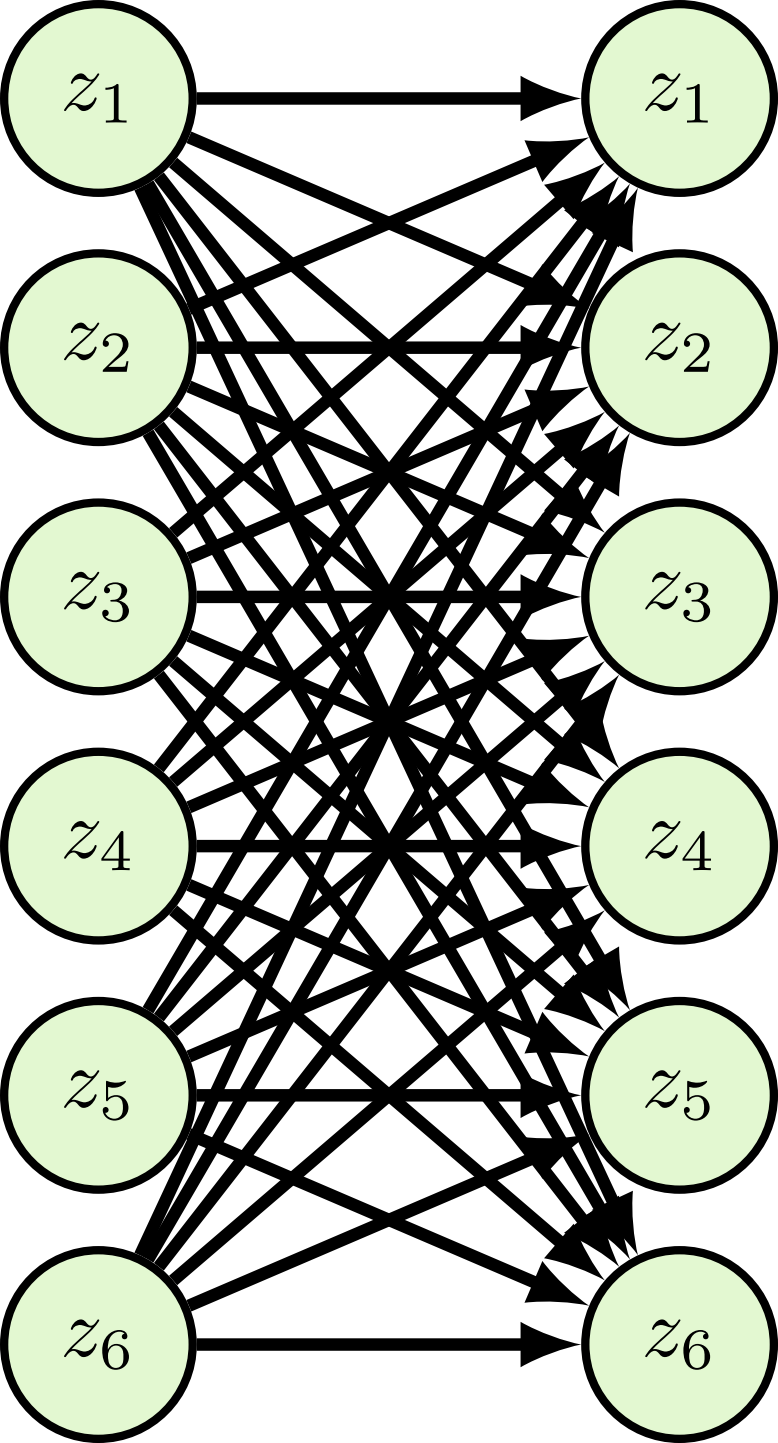

$f(x)$
x
1
-1
-3
-2
-1
1
2
3

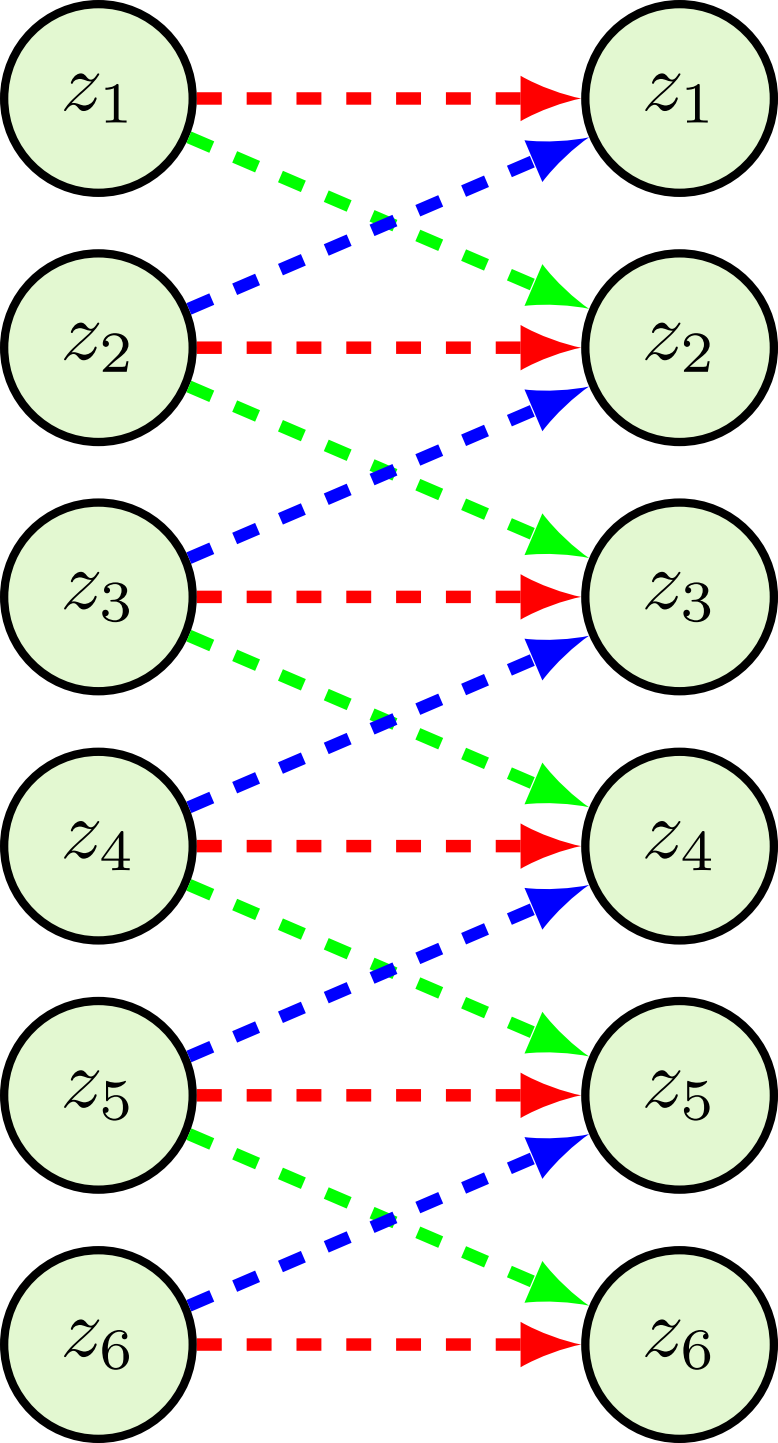

$f(x)$
x
1
-1
-3
-2
-1
1
2
3

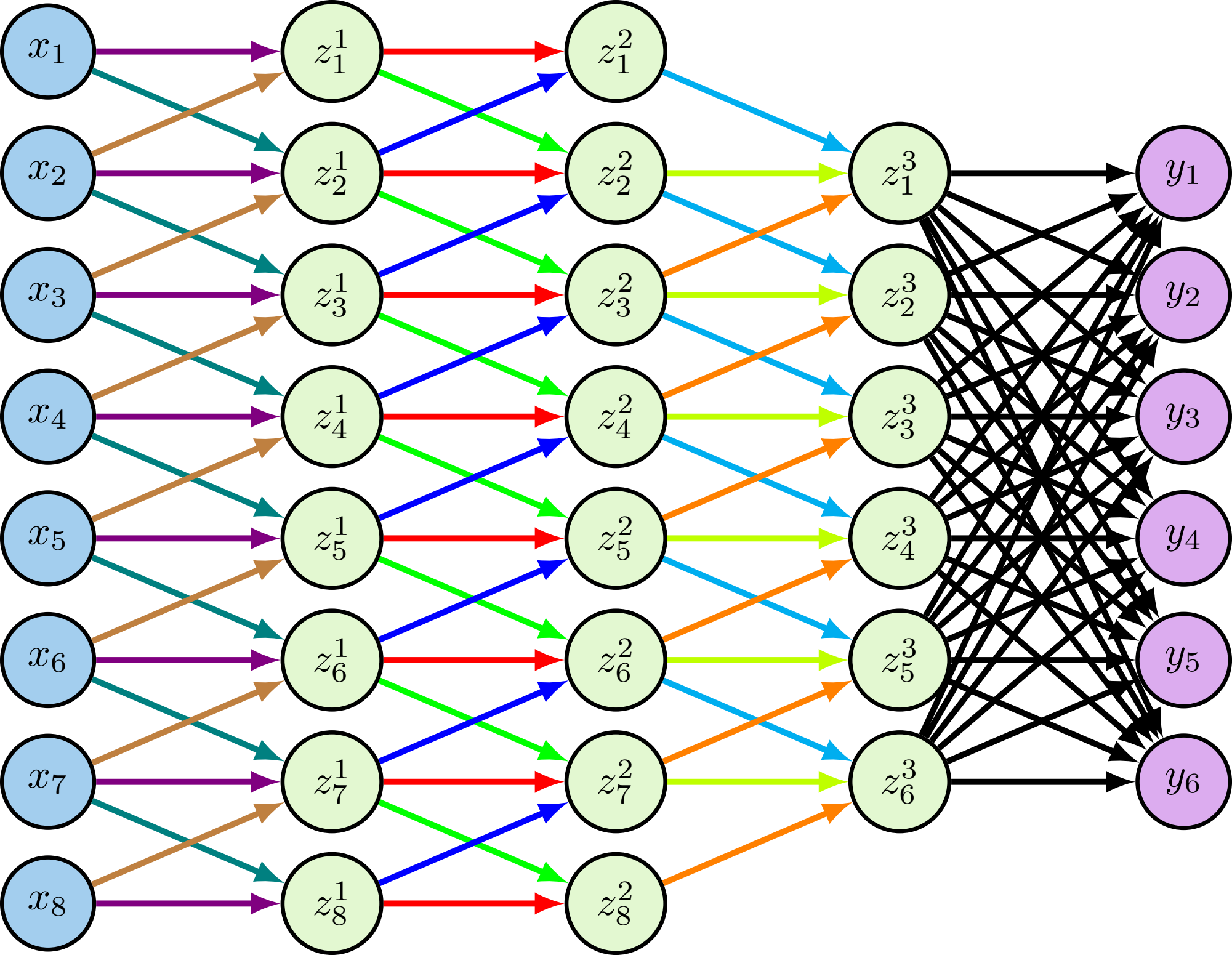

$f(x)$
x
1
-1
-3
-2
-1
1
2
3

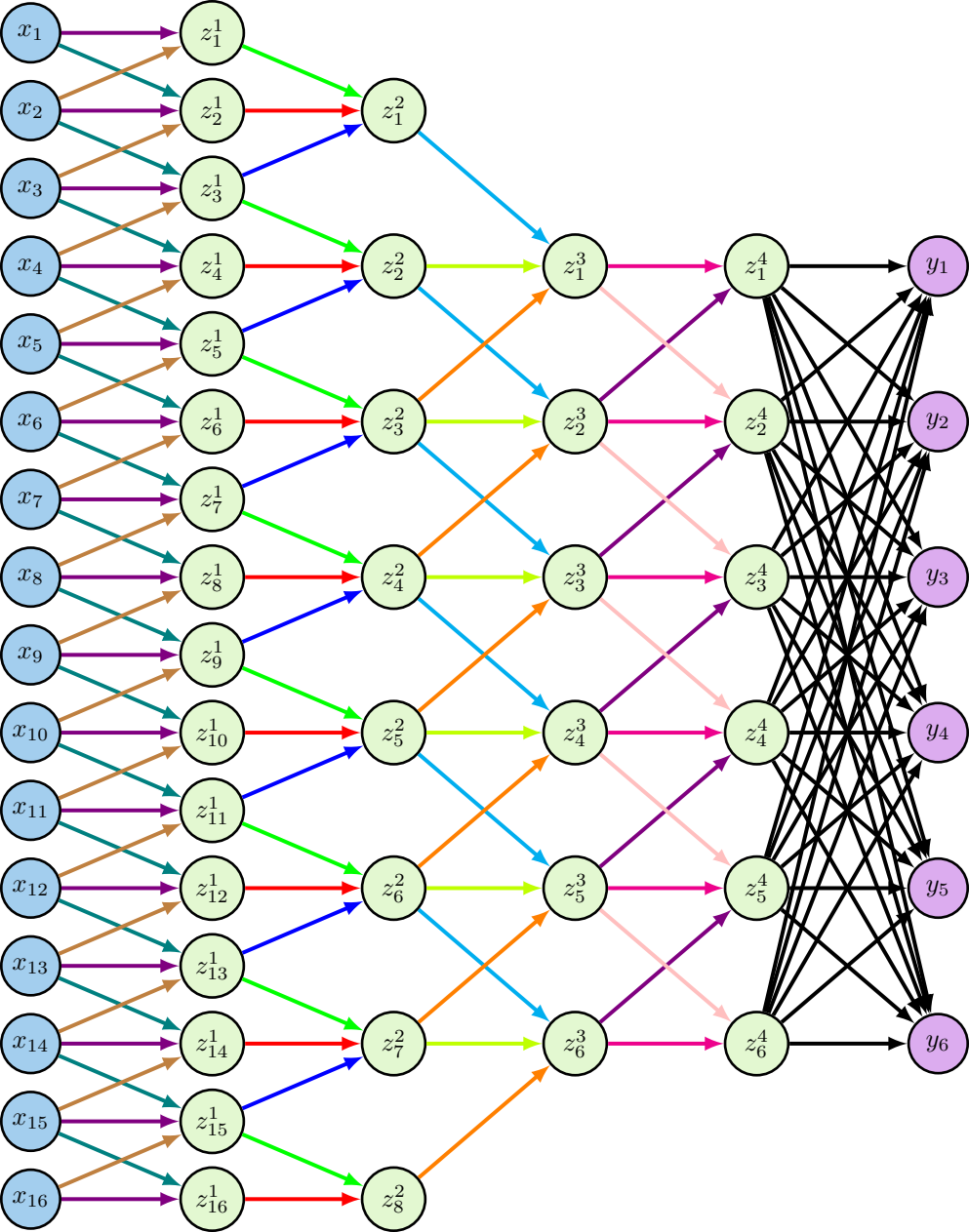

Input
layer
Hidden
layer
Output
layer
Input 1
Input 2
Input 3
x_1
x_2
x_3
z_1
z_2
z_3
z_4
y_1
Output 1

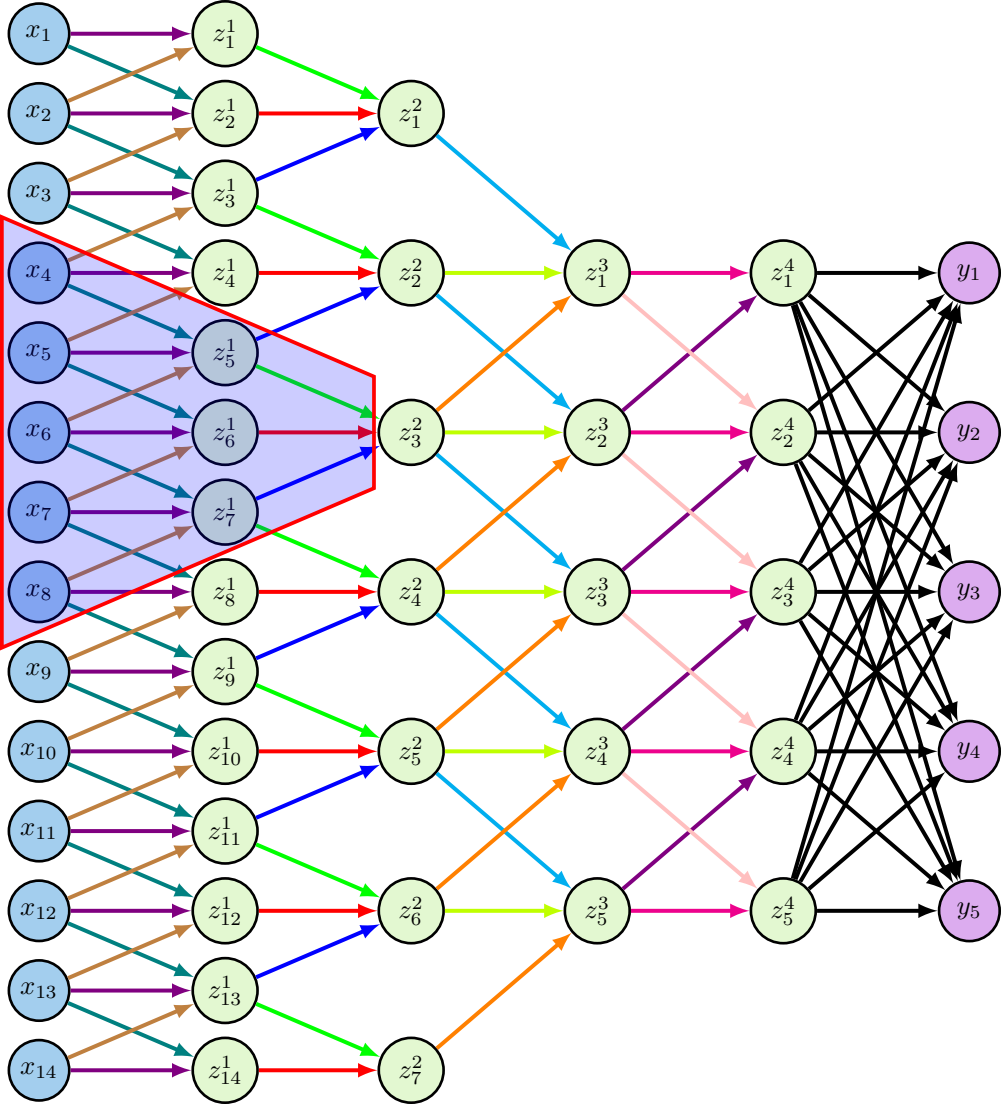

Input layer
Hidden layer 1
Hidden layer 2
Output layer
Input 1
Input 2
Input 3
x_1
x_2
x_3
z_1^1
z_2^1
z_3^1
z_4^1
z_1^2
z_2^2
z_3^2
z_4^2
y_1
y_2
Output 1
Output 2

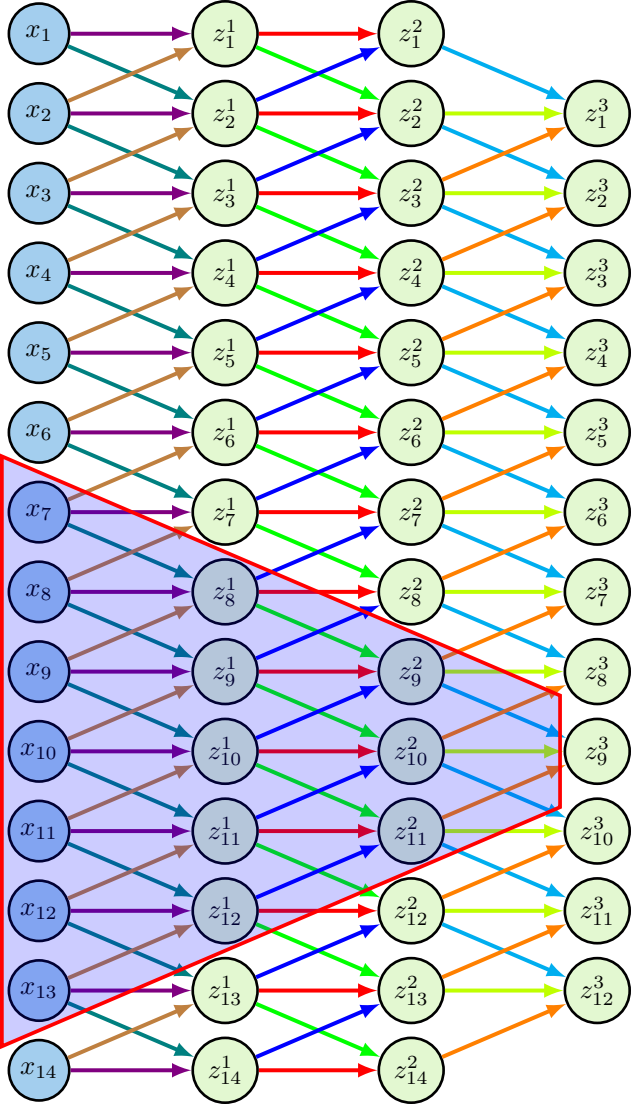

z_1
z_2
z_3
z_4
z_5
z_6
z_1
z_2
z_3
z_4
z_5
z_6

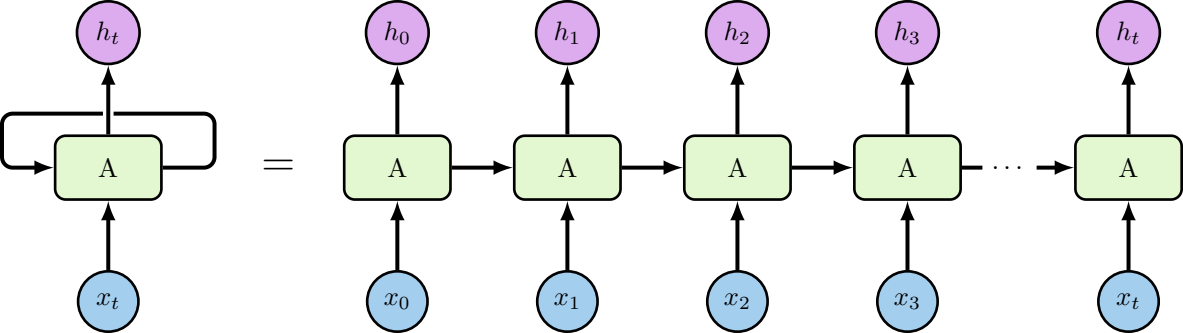

z_1
z_2
z_3
z_4
z_5
z_6
z_1
z_2
z_3
z_4
z_5
z_6

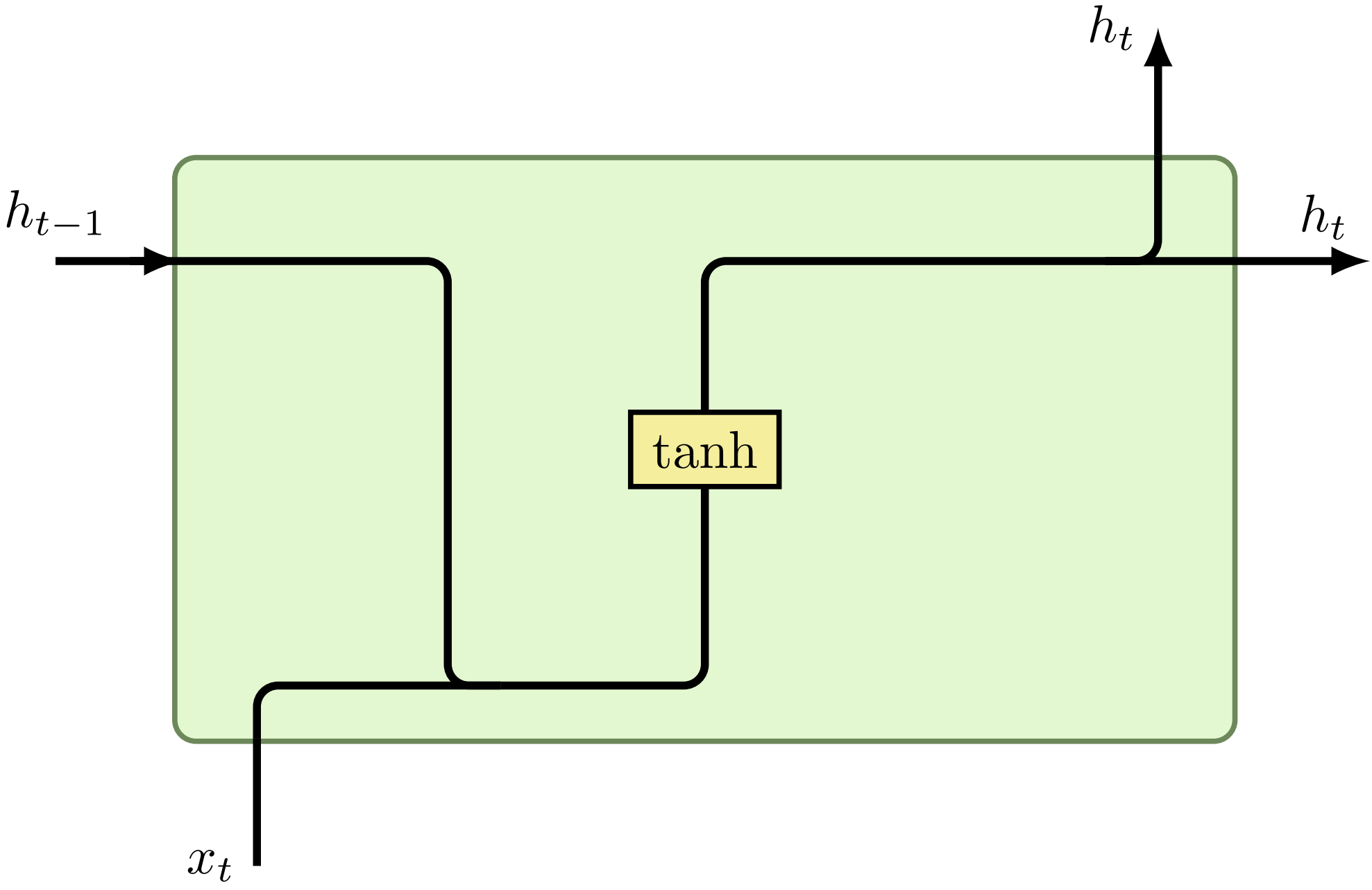

x_1
x_2
x_3
x_4
x_5
x_6
x_7
x_8
z_1^1
z_2^1
z_3^1
z_4^1
z_5^1
z_6^1
z_7^1
z_8^1
z_1^2
z_2^2
z_3^2
z_4^2
z_5^2
z_6^2
z_7^2
z_8^2
z_1^3
z_2^3
z_3^3
z_4^3
z_5^3
z_6^3
y_1
y_2
y_3
y_4
y_5
y_6

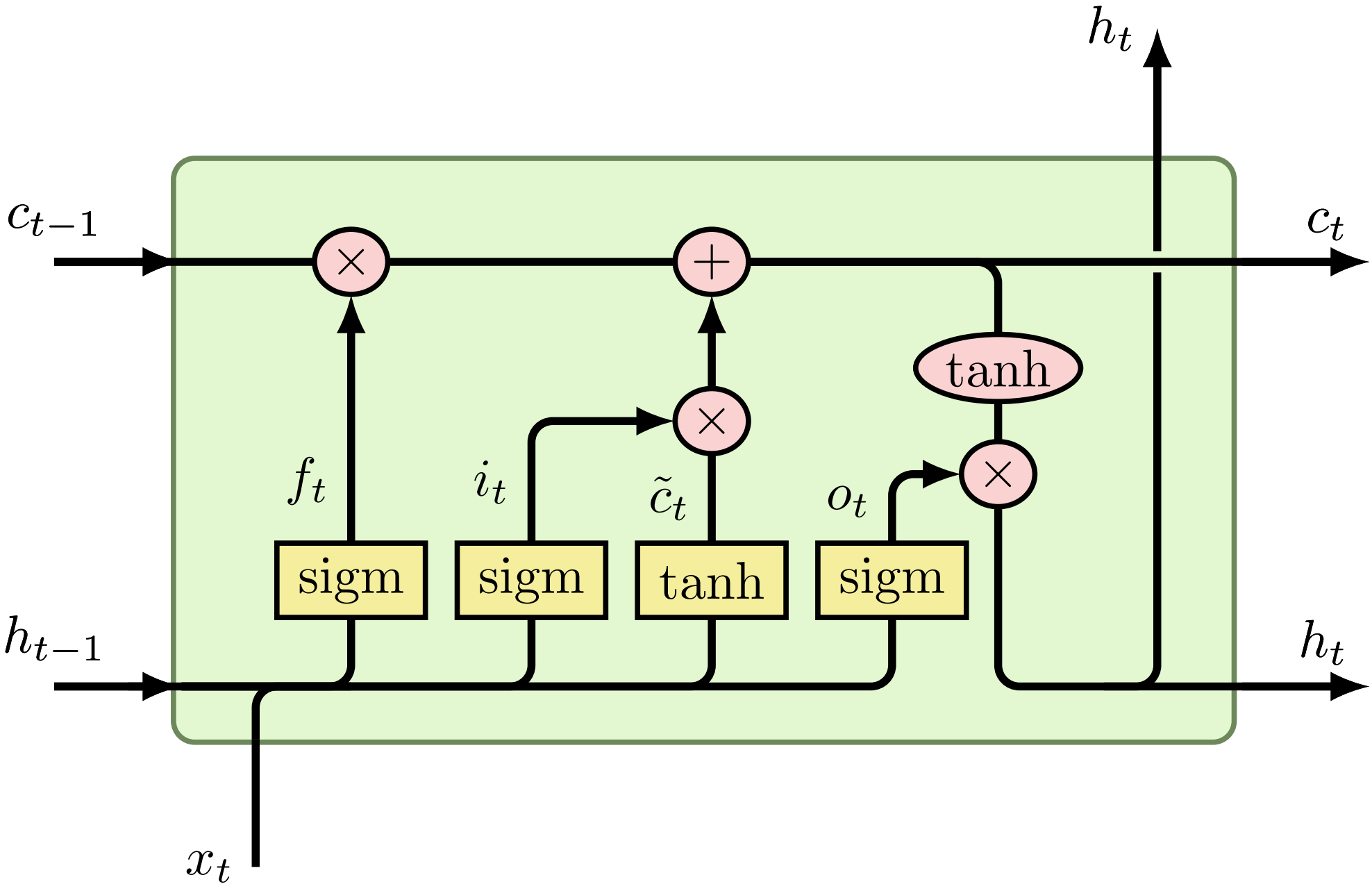

x_1
x_2
x_3
x_4
x_5
x_6
x_7
x_8
x_9
x_{10}
x_{11}
x_{12}
x_{13}
x_{14}
x_{15}
x_{16}
z_1^1
z_2^1
z_3^1
z_4^1
z_5^1
z_6^1
z_7^1
z_8^1
z_9^1
z_{10}^1
z_{11}^1
z_{12}^1
z_{13}^1
z_{14}^1
z_{15}^1
z_{16}^1
z_1^2
z_2^2
z_3^2
z_4^2
z_5^2
z_6^2
z_7^2
z_8^2
z_1^3
z_2^3
z_3^3
z_4^3
z_5^3
z_6^3
z_1^4
z_2^4
z_3^4
z_4^4
z_5^4
z_6^4
y_1
y_2
y_3
y_4
y_5
y_6

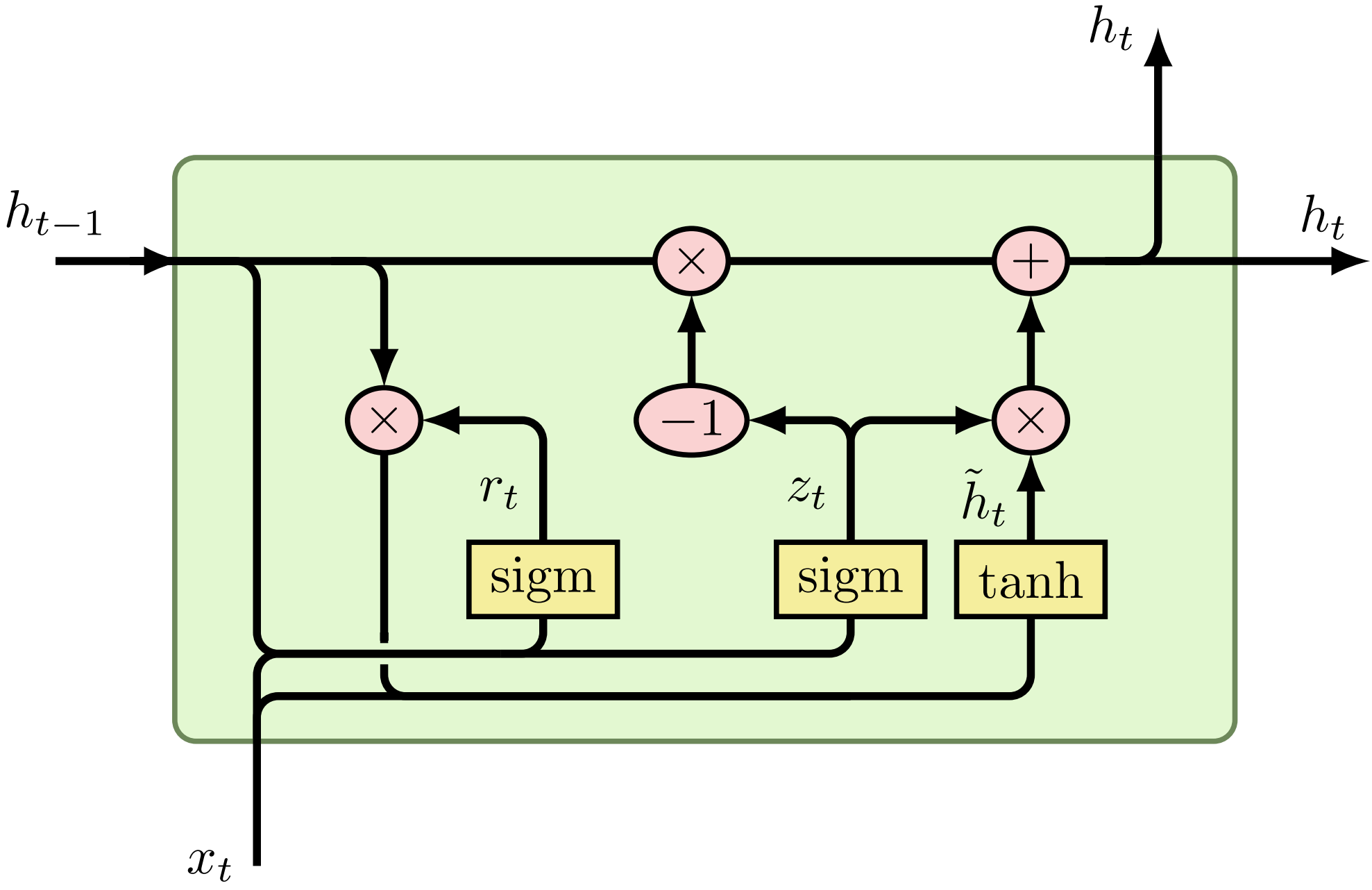

x_1
x_2
x_3
x_4
x_5
x_6
x_7
x_8
x_9
x_{10}
x_{11}
x_{12}
x_{13}
x_{14}
z_1^1
z_2^1
z_3^1
z_4^1
z_5^1
z_6^1
z_7^1
z_8^1
z_9^1
z_{10}^1
z_{11}^1
z_{12}^1
z_{13}^1
z_{14}^1
z_1^2
z_2^2
z_3^2
z_4^2
z_5^2
z_6^2
z_7^2
z_1^3
z_2^3
z_3^3
z_4^3
z_5^3
z_1^4
z_2^4
z_3^4
z_4^4
z_5^4
y_1
y_2
y_3
y_4
y_5


x_1
x_2
x_3
x_4
x_5
x_6
x_7
x_8
x_9
x_{10}
x_{11}
x_{12}
x_{13}
x_{14}
z_1^1 z_2^1 z_3^1 z_4^1 z_5^1 z_6^1 z_7^1 z_8^1 z_9^1 z_{10}^1 z_{11}^1 z_{12}^1 z_{13}^1 z_{14}^1
z_1^2 z_2^2 z_3^2 z_4^2 z_5^2 z_6^2 z_7^2 z_8^2 z_9^2 z_{10}^2 z_{11}^2 z_{12}^2 z_{13}^2 z_{14}^2
z_1^3 z_2^3 z_3^3 z_4^3 z_5^3 z_6^3 z_7^3 z_8^3 z_9^3 z_{10}^3 z_{11}^3 z_{12}^3

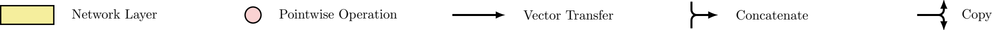

h_t
A
x_t
=
h_0
h_1
h_2
h_3
h_t
A
A
A
A
A
x_0
x_1
x_2
x_3
x_t

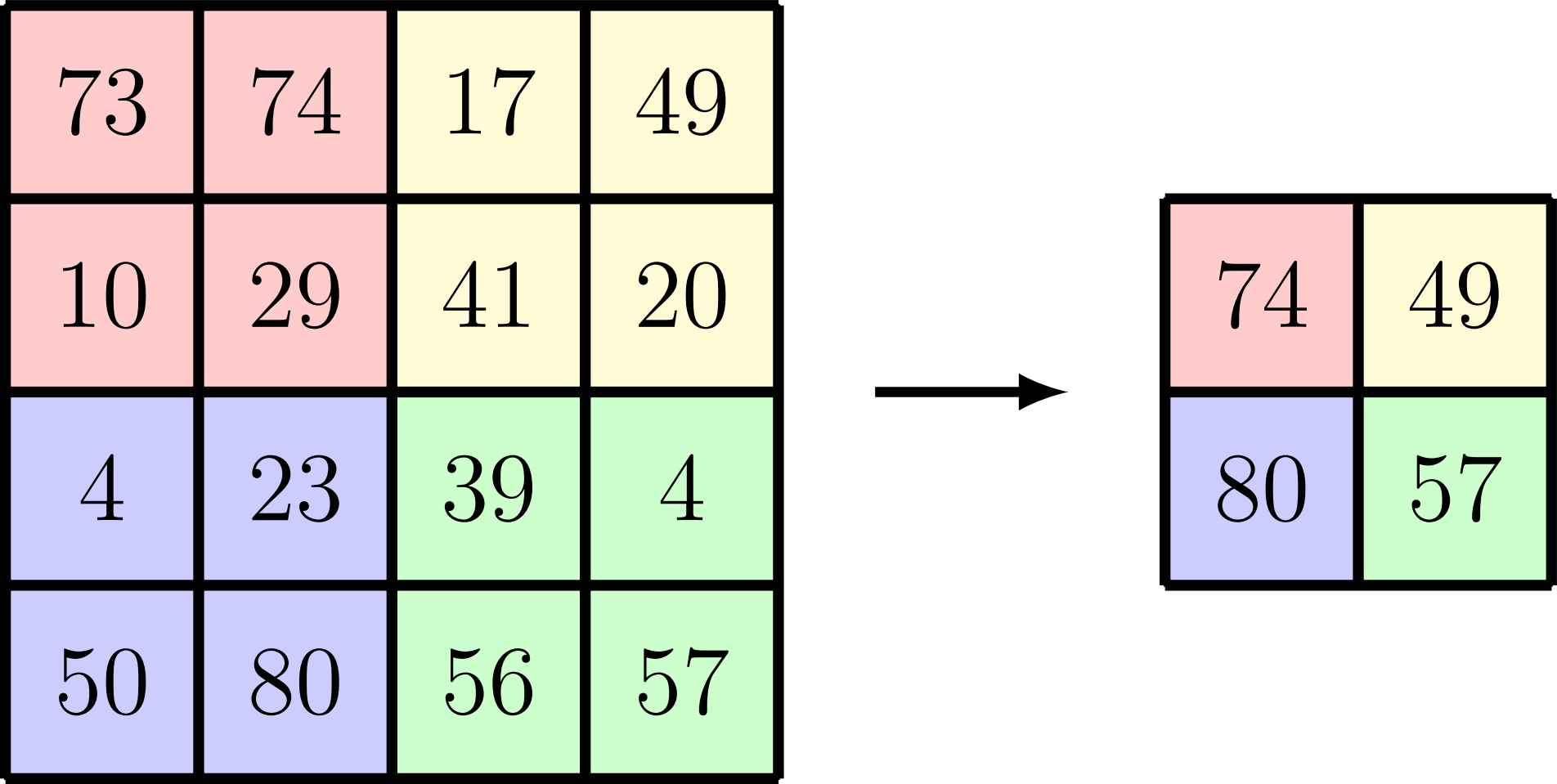

h_t
h_{t-1}
h_t
tanh
x_t

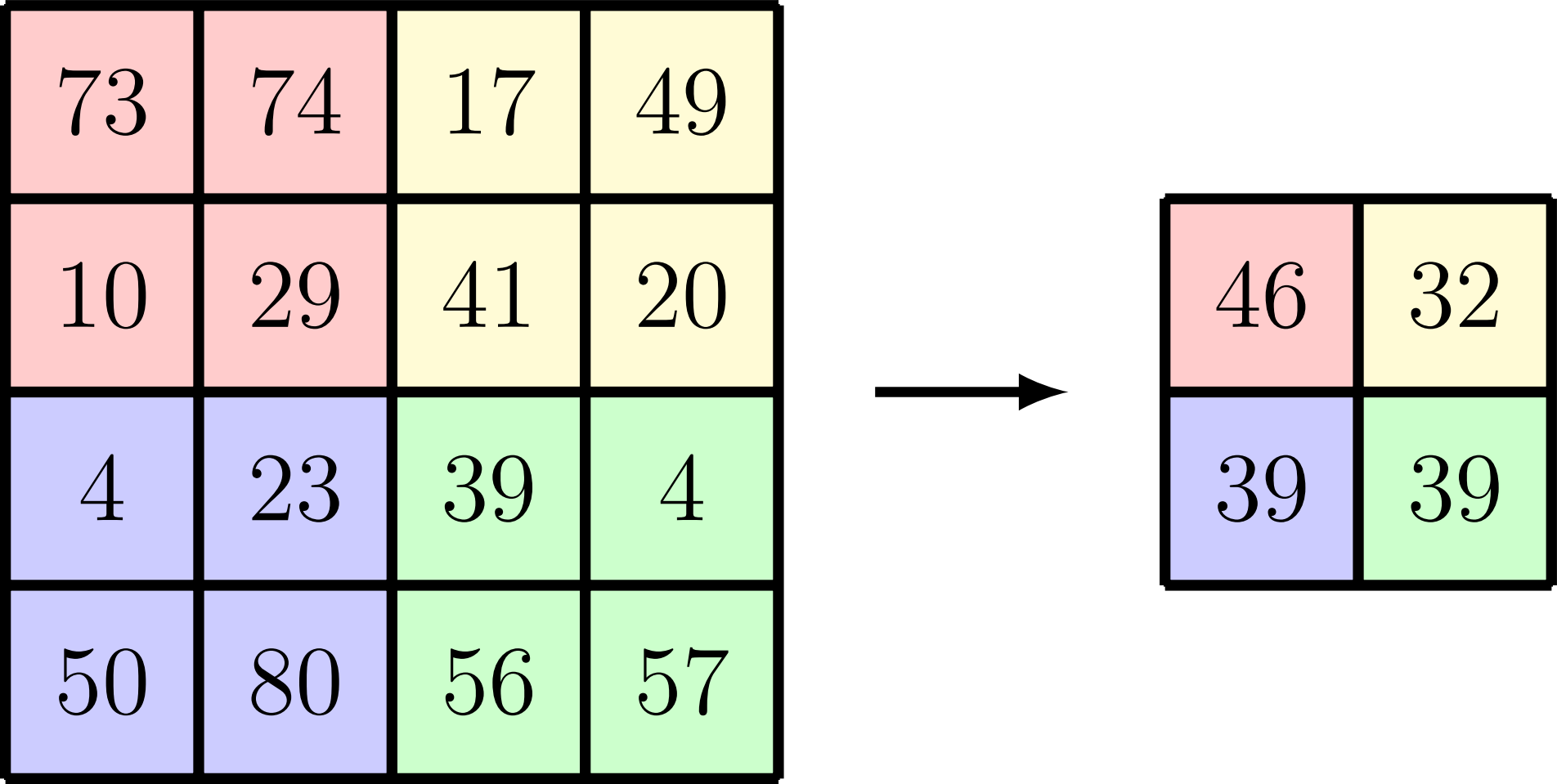

h_t
c_{t-1}
c_t
tanh
f_t
i_t
$\tilde{c}_t$
o_t
sigm
sigm
tanh
sigm
h_{t-1}
h_t
x_t


h_t
h_{t-1}
h_t
-1
r_t
z_t
$\tilde{h}_t$
sigm
sigm
tanh
x_t


Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

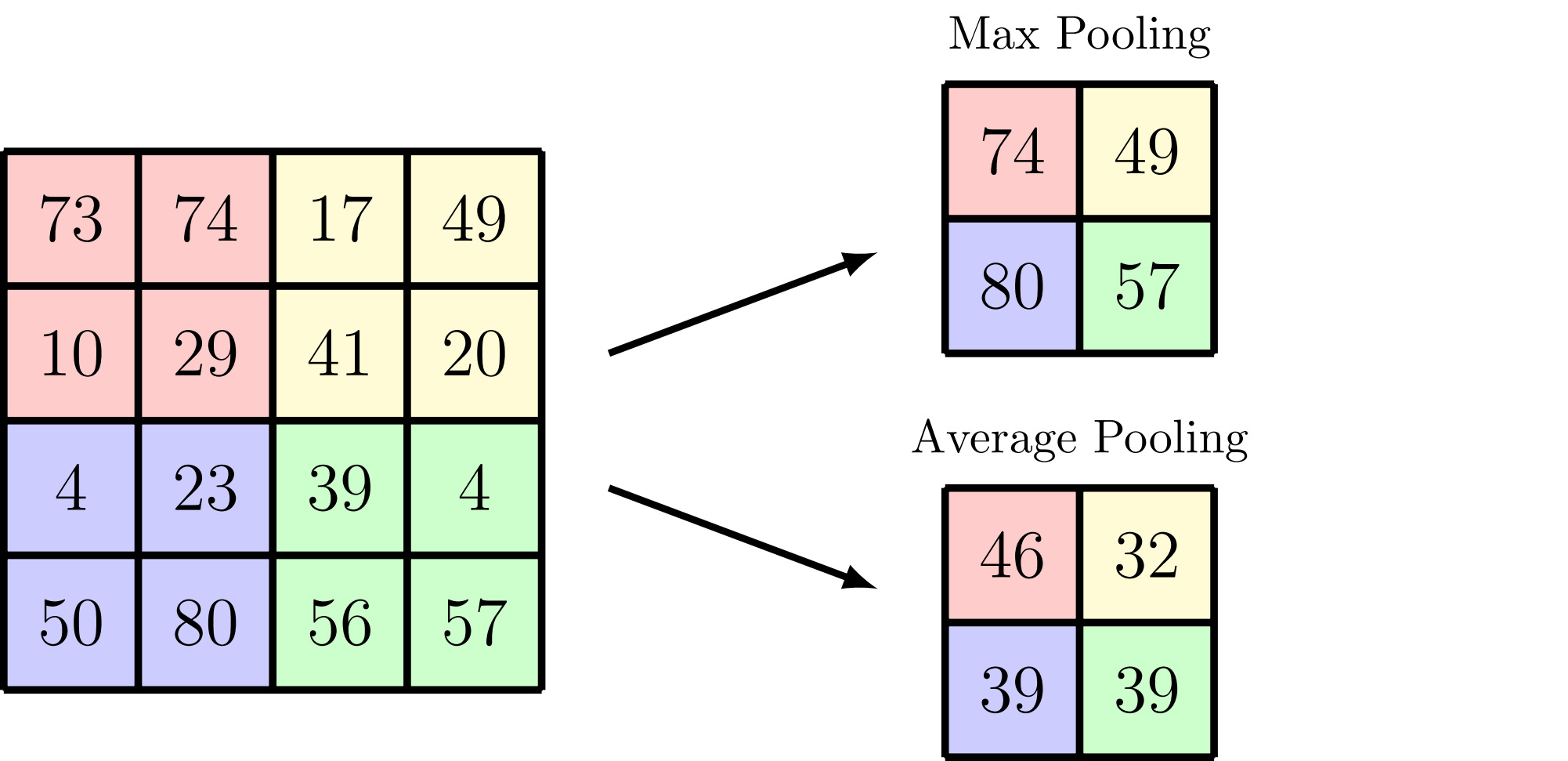

Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

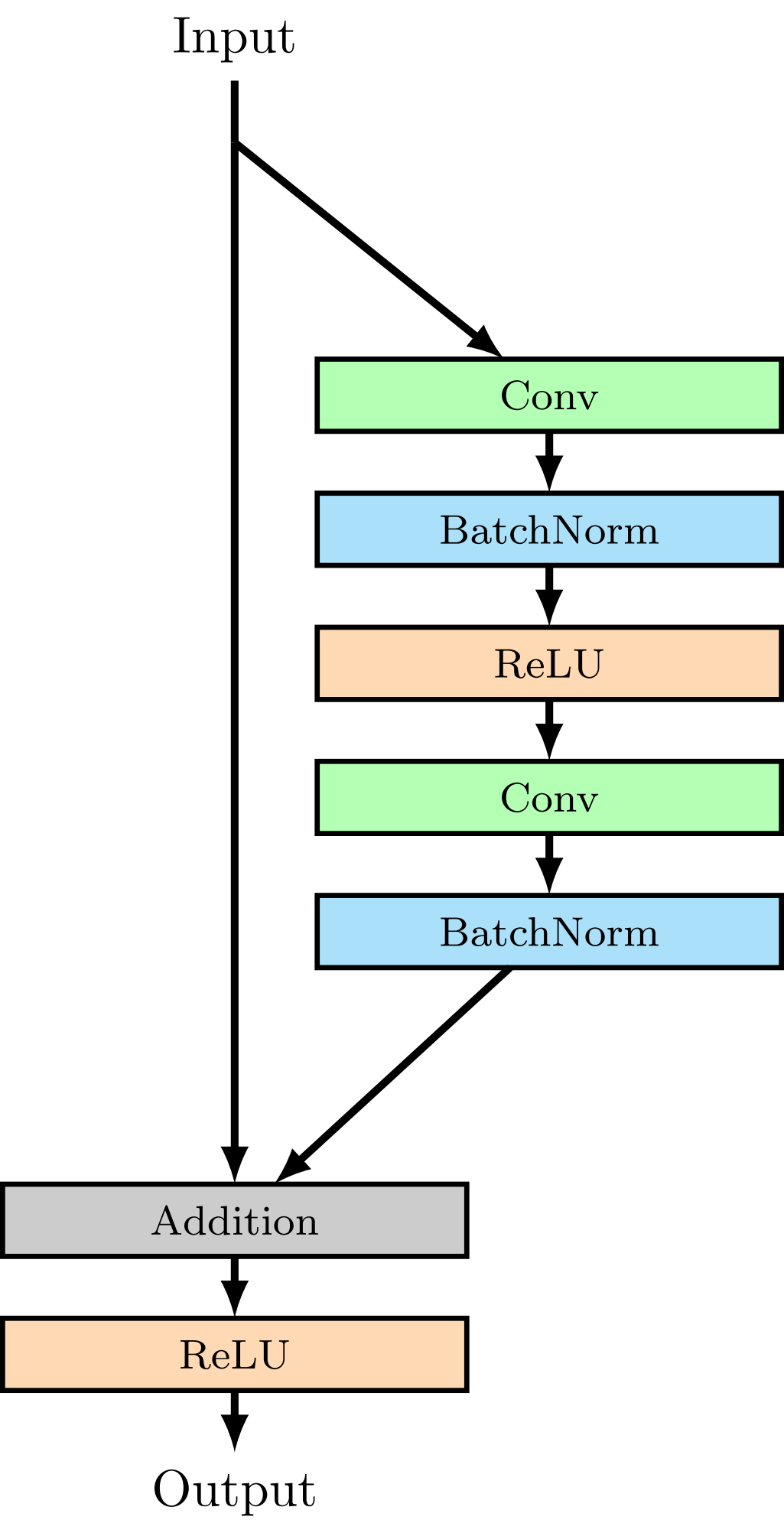

73 74 17 49
10 29 41 20
4 23 39 4
50 80 56 57
74 49
80 57

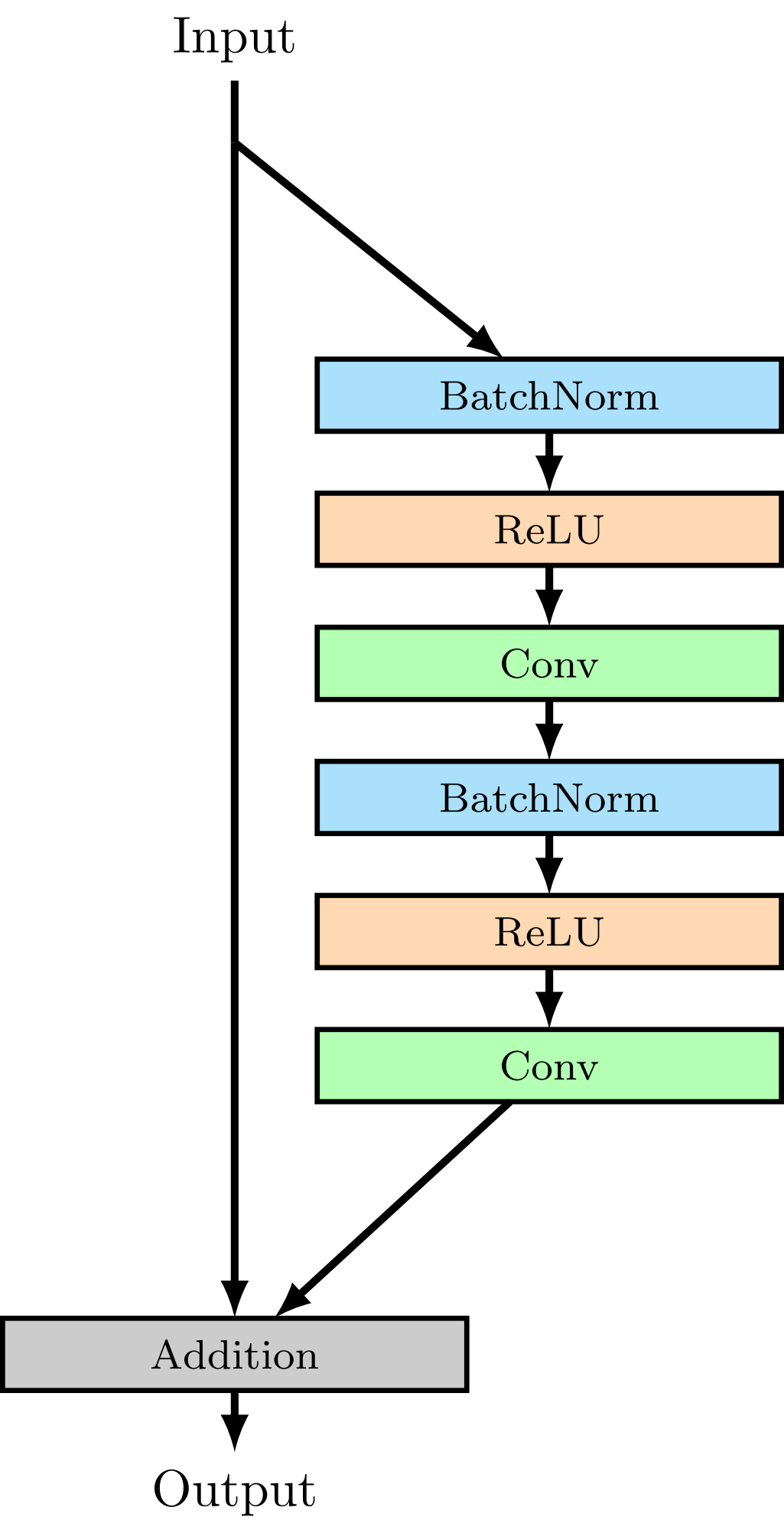

73 74 17 49
10 29 41 20
4 23 39 4
50 80 56 57
46 32
39 39

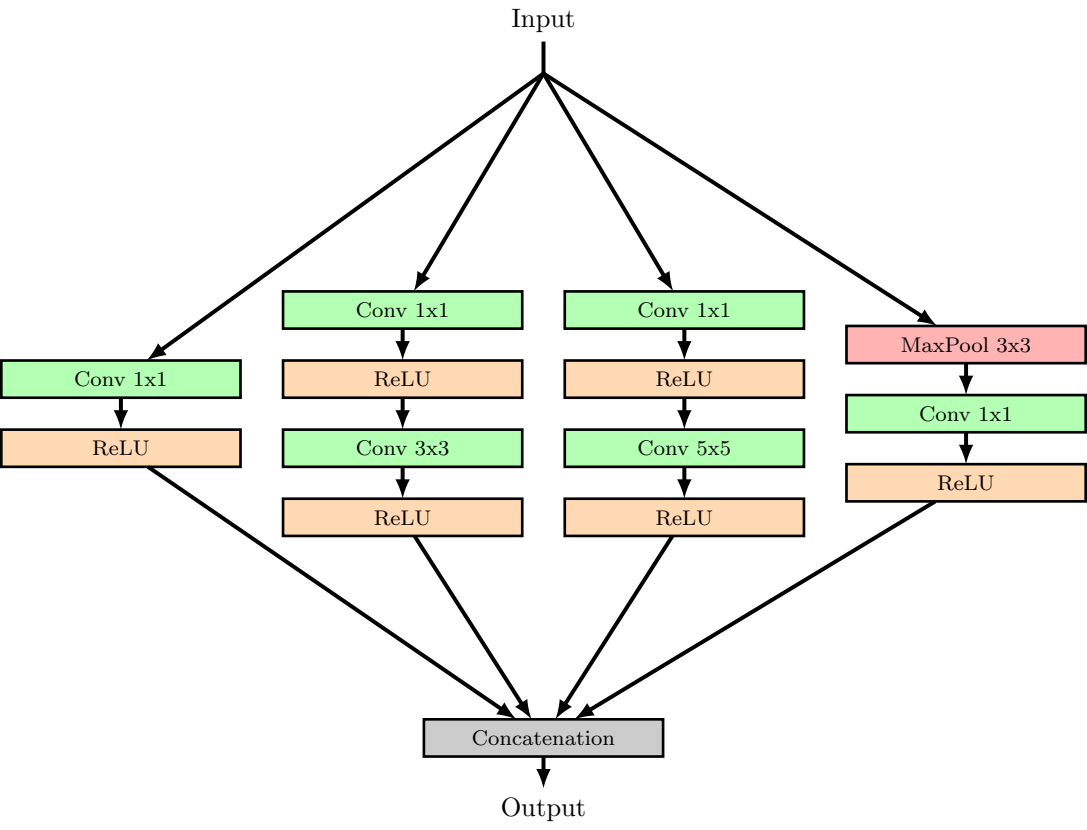
90	15	19	36	85
13	55	31	42	23
41	11	11	57	58
67	89	87	83	49
2	11	19	32	91

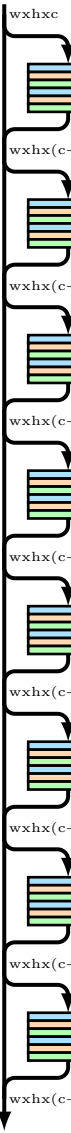
90	85
89	91

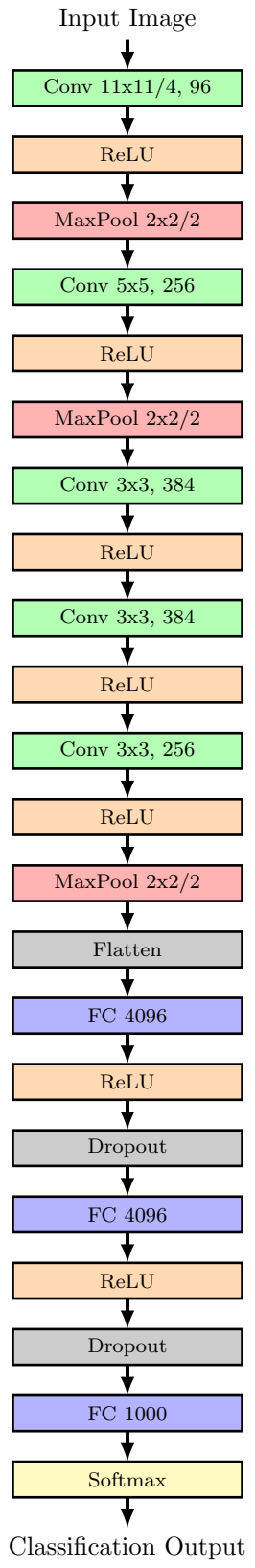

73 74 17 49
10 29 41 20
4 23 39 4
50 80 56 57
Max Pooling
74 49
80 57
Average Pooling
46 32
39 39


Input
Conv
BatchNorm
ReLU
Conv
BatchNorm
Addition
ReLU
Output

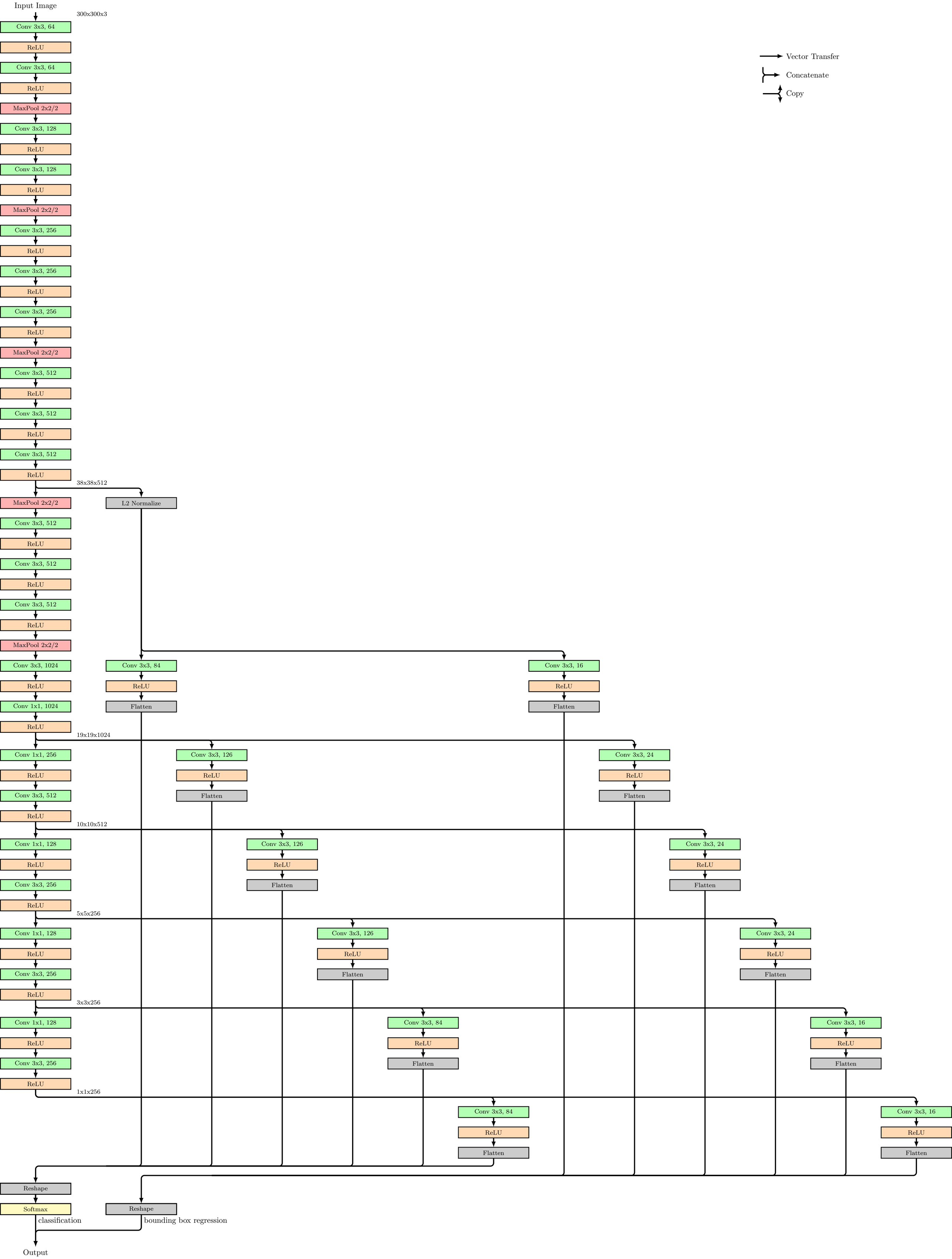

Input
BatchNorm
ReLU
Conv
BatchNorm
ReLU
Conv
Addition
Output

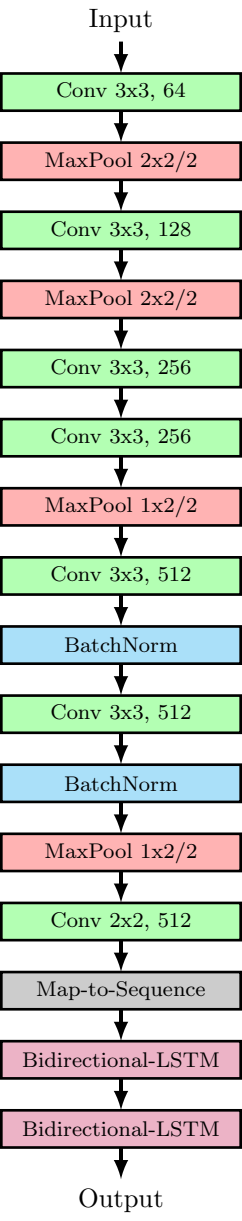

Input
Conv 1x1
ReLU
Conv 1x1
ReLU
Conv 3x3
ReLU
Conv 1x1
ReLU
Conv 5x5
ReLU
MaxPool 3x3
Conv 1x1
ReLU
Concatenation
Output

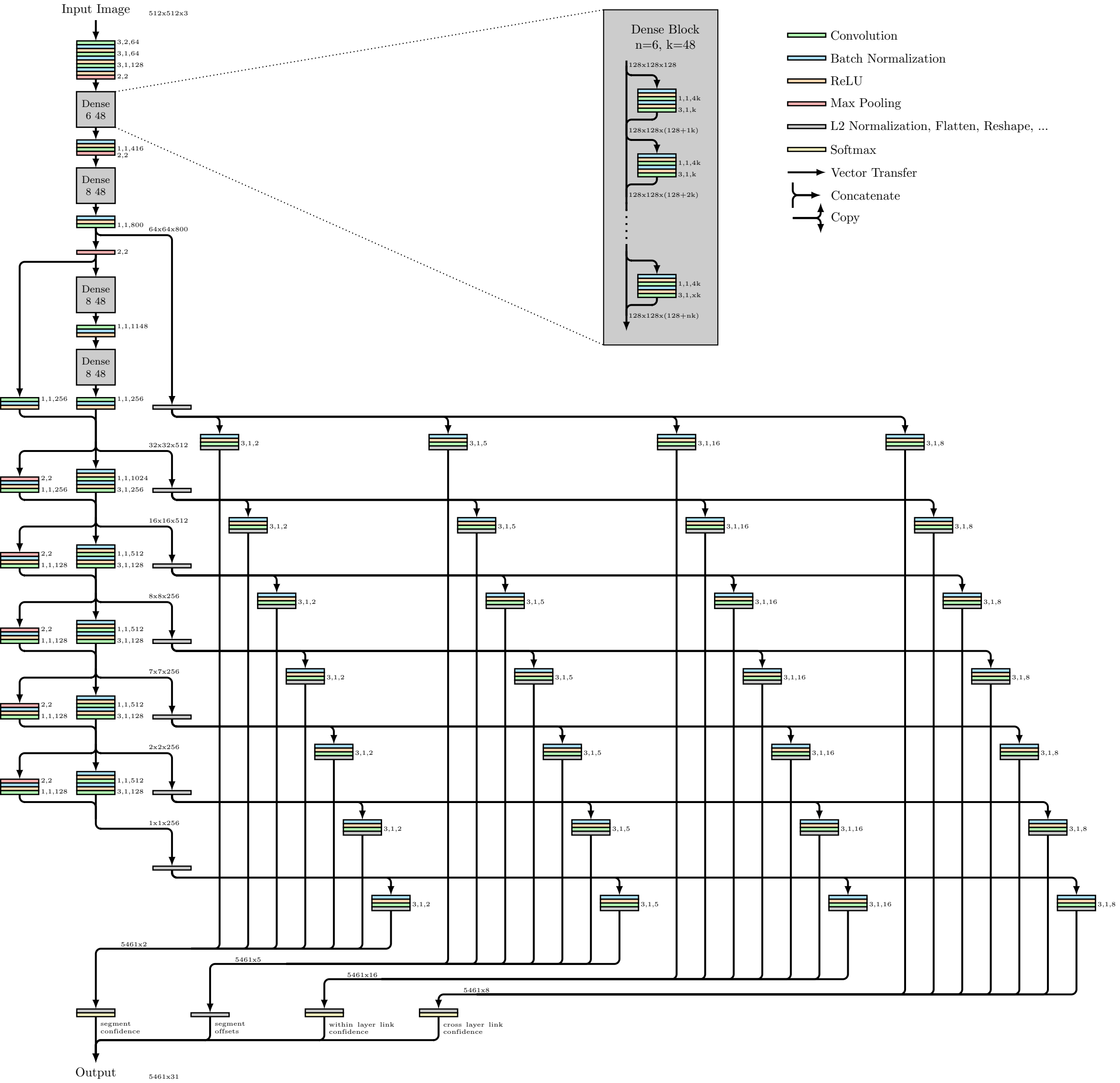

wxhxc
1,1,4k
3,1,k
wxhx(c+1k)
1,1,4k
3,1,k
wxhx(c+2k)
1,1,4k
3,1,k
wxhx(c+3k)
1,1,4k
3,1,k
wxhx(c+4k)
1,1,4k
3,1,k
wxhx(c+5k)
1,1,4k
3,1,k
wxhx(c+6k)
1,1,4k
3,1,k
wxhx(c+7k)
1,1,4k
3,1,k
wxhx(c+8k)

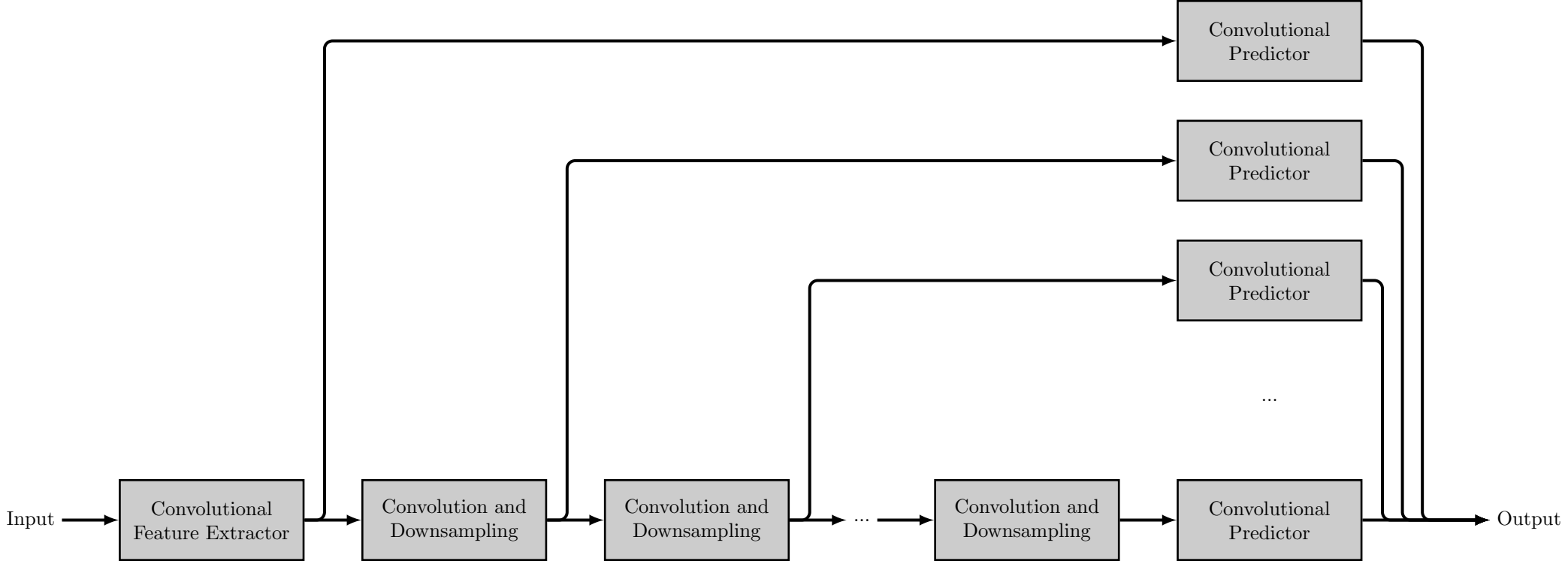

Input Image
Conv 11x11/4, 96
ReLU
MaxPool 2x2/2
Conv 5x5, 256
ReLU
MaxPool 2x2/2
Conv 3x3, 384
ReLU
Conv 3x3, 384
ReLU
Conv 3x3, 256
ReLU
MaxPool 2x2/2
Flatten
FC 4096
ReLU
Dropout
FC 4096
ReLU
Dropout
FC 1000
Softmax
Classification Output

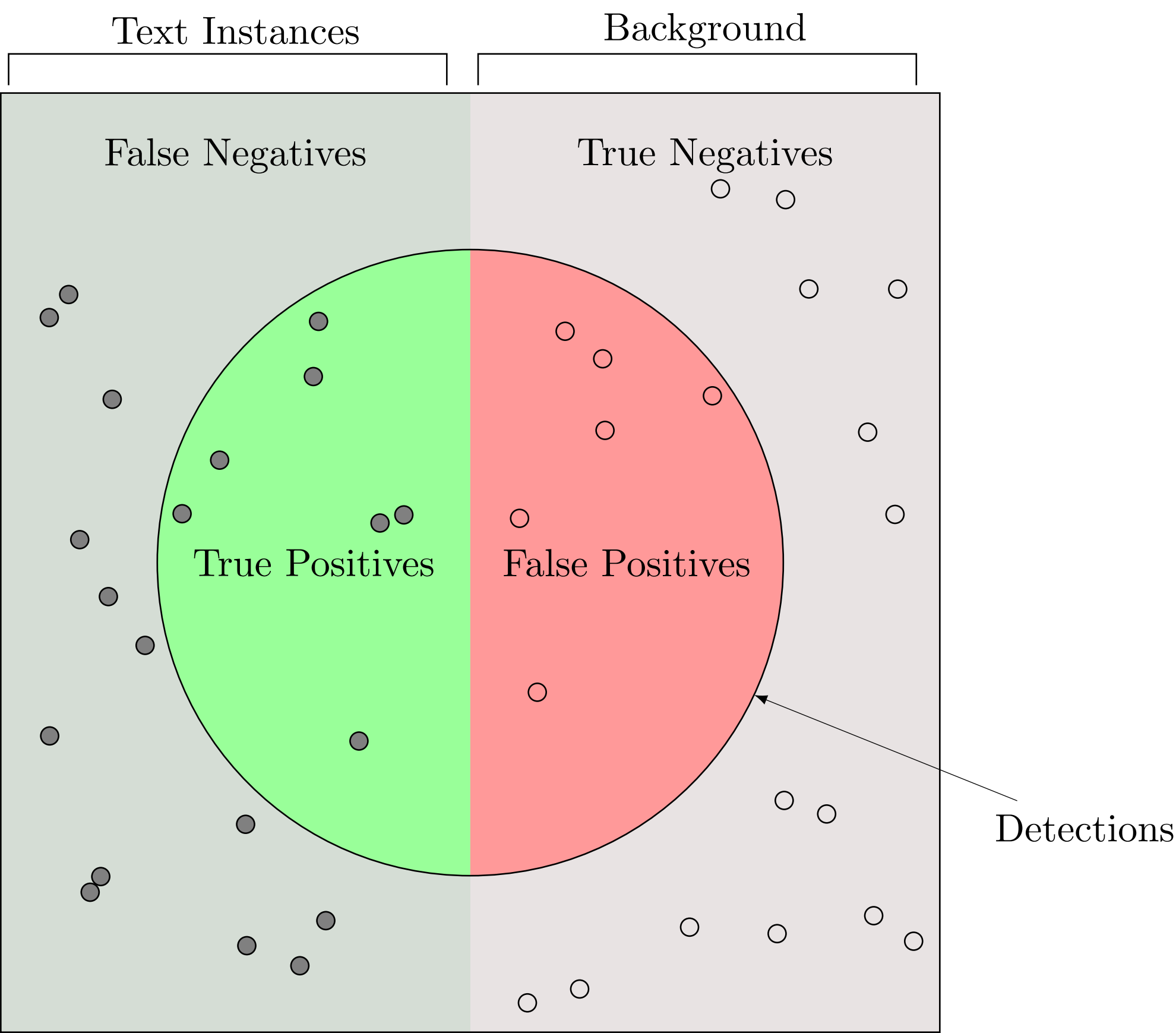

Input Image
Conv 3x3, 64
ReLU
Conv 3x3, 64
ReLU
MaxPool 2x2/2
Conv 3x3, 64
ReLU
Conv 3x3, 64
ReLU
MaxPool 2x2/2
Conv 3x3, 256
ReLU
Conv 3x3, 256
ReLU
Conv 3x3, 256
ReLU
MaxPool 2x2/2
Conv 3x3, 512
ReLU
Conv 3x3, 512
ReLU
MaxPool 2x2/2
Conv 3x3, 512
ReLU
Conv 3x3, 512
ReLU
Conv 3x3, 512
ReLU
MaxPool 2x2/2
Flatten
FC 4096
ReLU
Dropout
FC 4096
ReLU
Dropout
FC 1000
Softmax
Classification Output


Input Image
300x300x3
Conv 3x3, 64
ReLU
Conv 3x3, 64
ReLU
MaxPool 2x2/2
Conv 3x3, 128
ReLU
Conv 3x3, 128
ReLU
MaxPool 2x2/2
Conv 3x3, 256
ReLU
Conv 3x3, 256
ReLU
Conv 3x3, 256
ReLU
MaxPool 2x2/2
Conv 3x3, 512
ReLU
Conv 3x3, 512
ReLU
Conv 3x3, 512
ReLU
38x38x512
MaxPool 2x2/2
L2 Normalize
Conv 3x3, 512
ReLU
Conv 3x3, 512
ReLU
Conv 3x3, 512
ReLU
MaxPool 2x2/2
Conv 3x3, 1024
ReLU
Conv 1x1, 1024
ReLU
Conv 3x3, 84
ReLU
Flatten
Conv 3x3, 16
ReLU
Flatten
19x19x1024
Conv 1x1, 256
ReLU
Conv 3x3, 512
ReLU
Conv 3x3, 126
ReLU
Flatten
Conv 3x3, 24
ReLU
Flatten
10x10x512
Conv 1x1, 128
ReLU
Conv 3x3, 256
ReLU
Conv 3x3, 126
ReLU
Flatten
Conv 3x3, 24
ReLU
Flatten
5x5x256
Conv 1x1, 128
ReLU
Conv 3x3, 256
ReLU
Conv 3x3, 126
ReLU
Flatten
Conv 3x3, 24
ReLU
Flatten
3x3x256
Conv 1x1, 128
ReLU
Conv 3x3, 256
ReLU
Conv 3x3, 84
ReLU
Flatten
Conv 3x3, 16
ReLU
Flatten
1x1x256
Conv 3x3, 84
ReLU
Flatten
Conv 3x3, 16
ReLU
Flatten
Reshape
Softmax
classification
Reshape
bounding box regression
Output
Vector Transfer
Concatenate
Copy


Input
Conv 3x3, 64
MaxPool 2x2/2
Conv 3x3, 128
MaxPool 2x2/2
Conv 3x3, 256
Conv 3x3, 256
MaxPool 1x2/2
Conv 3x3, 512
BatchNorm
Conv 3x3, 512
BatchNorm
MaxPool 1x2/2
Conv 2x2, 512
Map-to-Sequence
Bidirectional-LSTM
Bidirectional-LSTM
Output


Input Image
512x512x3
3,2,64
3,1,64
3,1,128
2,2
Dense
6 48
1,1,416
2,2
Dense
8 48
1,1,800
64x64x800
2,2
Dense
8 48
1,1,1148
Dense
8 48
1,1,256
1,1,256
32x32x512
2,2
1,1,256
1,1,1024
3,1,256
16x16x512
2,2
1,1,128
1,1,512
3,1,128
8x8x256
2,2
1,1,128
1,1,512
3,1,128
7x7x256
2,2
1,1,128
1,1,512
3,1,128
2x2x256
2,2
1,1,128
1,1,512
3,1,128
1x1x256
3,1,2
3,1,5
3,1,16
3,1,8
5461x2
5461x5
5461x16
5461x8
segment
confidence
segment
offsets
within layer link
confidence
cross layer link
confidence
Output
5461x31
Dense Block
n=6, k=48
128x128x128
1,1,4k
3,1,k
128x128x(128+1k)
1,1,4k
3,1,k
128x128x(128+2k)
1,1,4k
3,1,xk
128x128x(128+nk)
Convolution
Batch Normalization
ReLU
Max Pooling
L2 Normalization, Flatten, Reshape, ...
Softmax
Vector Transfer
Concatenate
Copy


Input
Convolutional Feature Extractor
Convolution and Downsampling
Convolution and Downsampling
...
Convolution and Downsampling
Convolutional Predictor
Convolutional Predictor
Convolutional Predictor
...
Convolutional Predictor
Output


Text Instances
Background
False Negatives
True Negatives
True Positives
False Positives
Detections