ConvNet

Ex. VGG16, ZF

Feature map

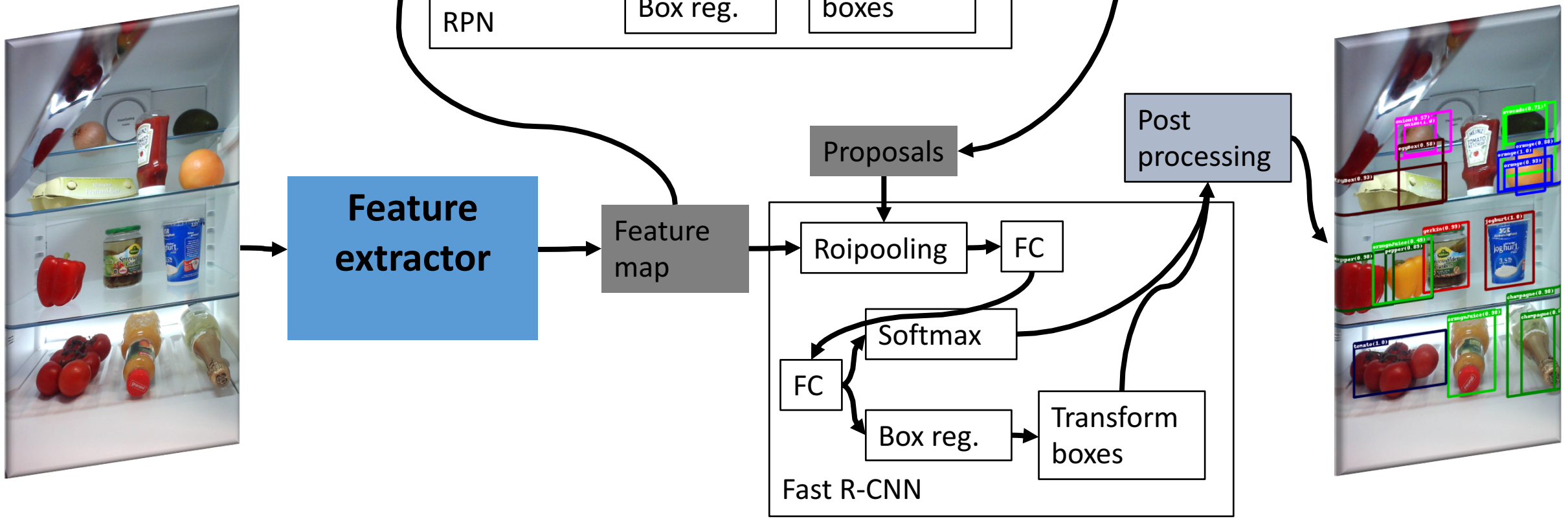
Proposals

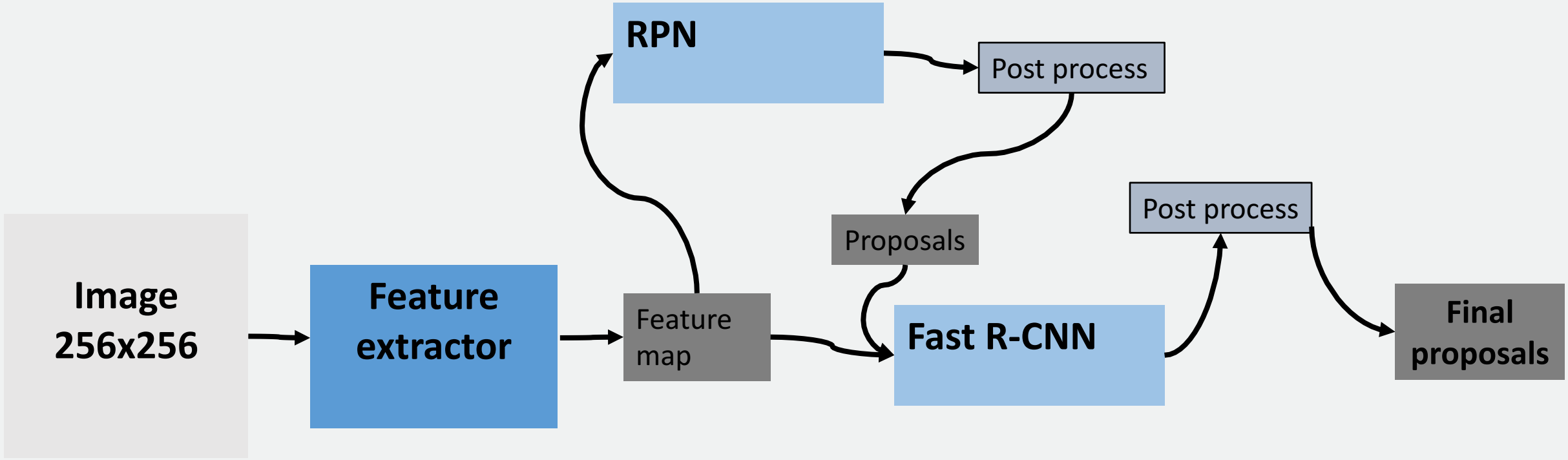
Classifier

Class 1

Class 2

Class 3


3x3 conv.
1x1 conv
Obj. det.
1x1 conv
Box reg.
Transform
boxes
RPN
Post
processing
Proposals
Feature
extractor
Feature
map
Roipooling
FC
Softmax
FC
Box reg.
Transform
boxes
Fast R-CNN
Post
processing


RPN
Post process
Image
256x256
Feature
extractor
Feature
map
Proposals
Fast R-CNN
Post process
Final
proposals

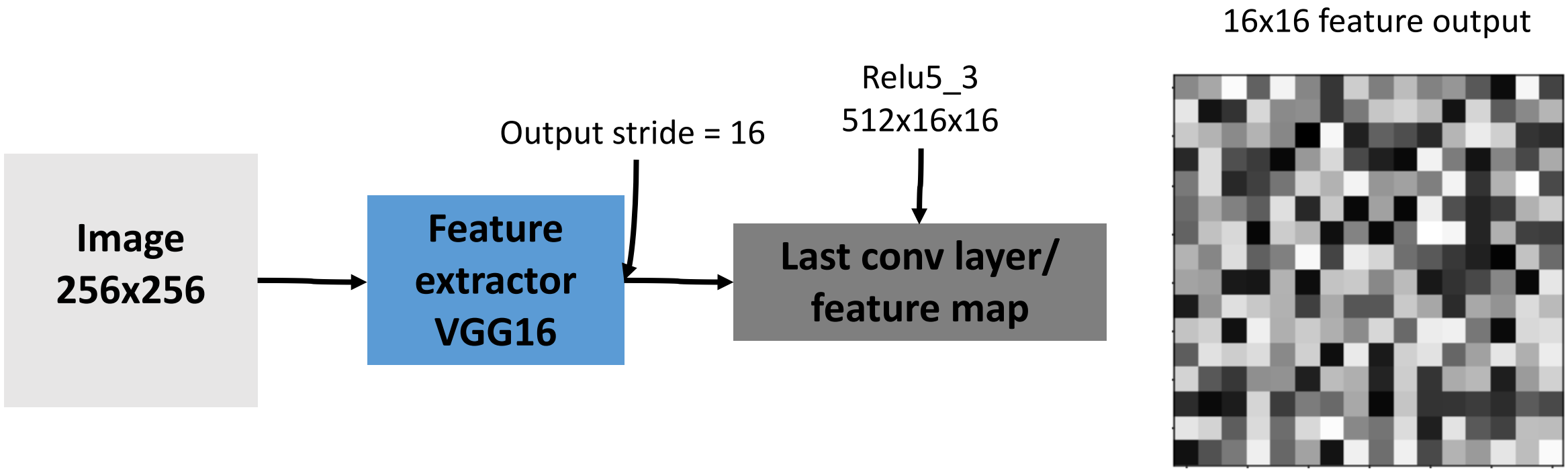
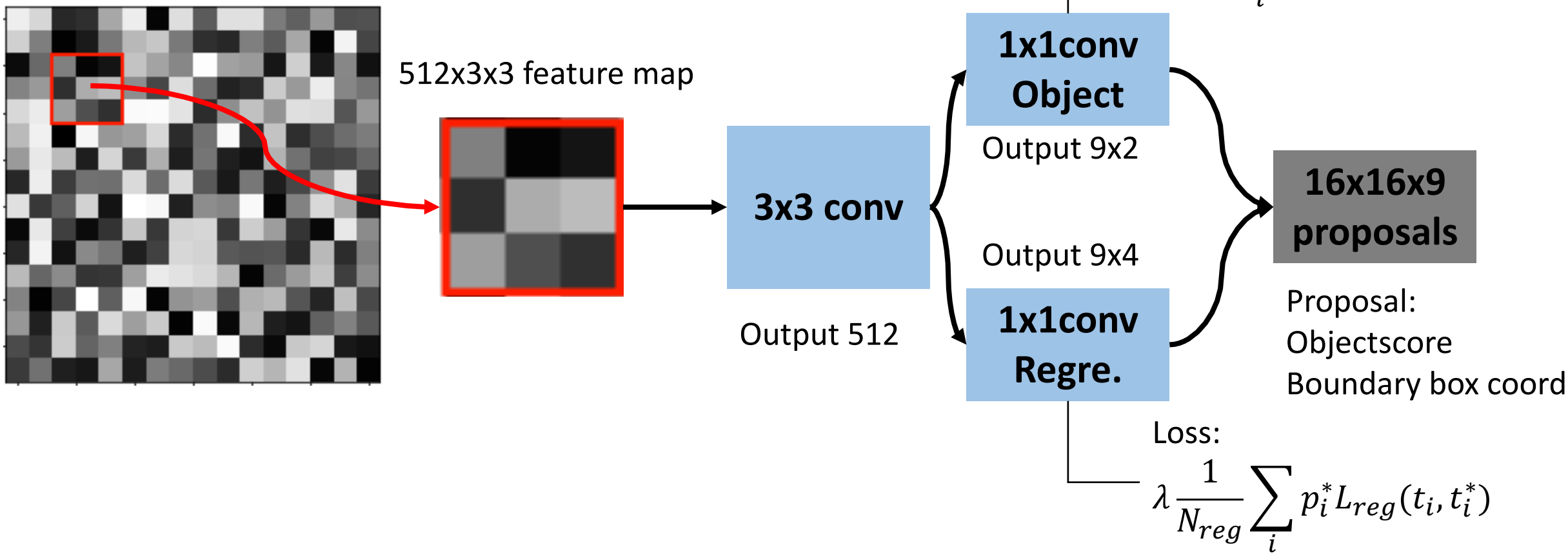
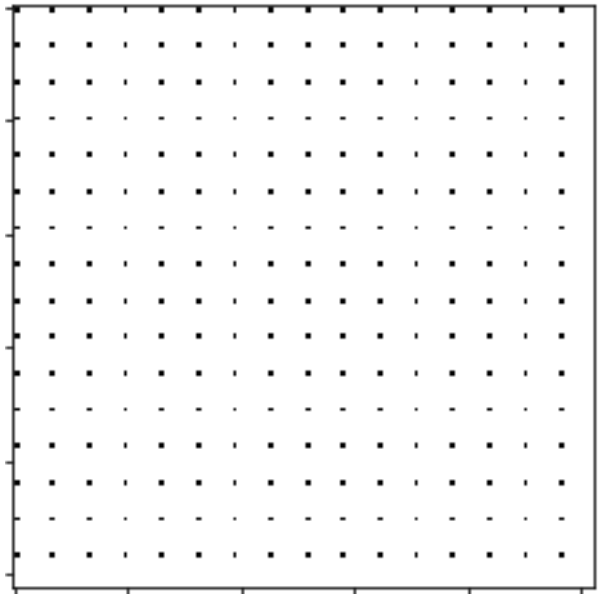


Image
256x256
Feature
extractor
VGG16
Output stride = 16
Relu5_3
512x16x16
Last conv layer/
feature map
16x16 feature output

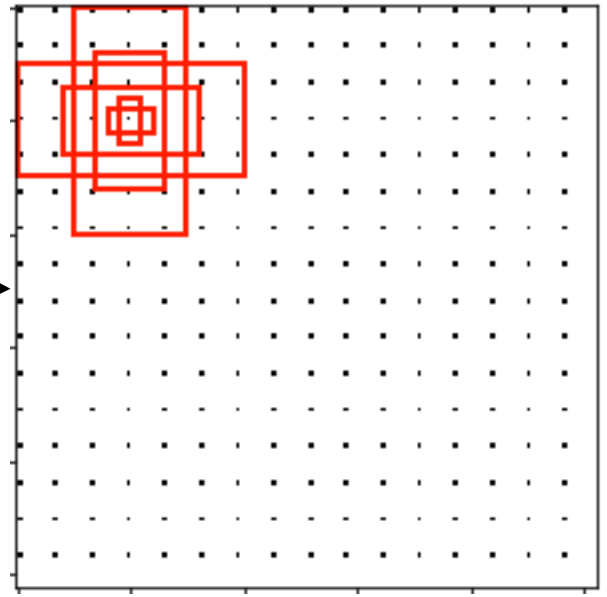

3x3 slidingwindow
512x3x3 feature map
3x3 conv
Output 512
1x1conv Object
Output 9x2
1x1conv Regre.
Output 9x4
16x16x9 proposals
Proposal:
Objectscore
Boundary box coord
Loss:
$\frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*)$
Loss:
$\lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$

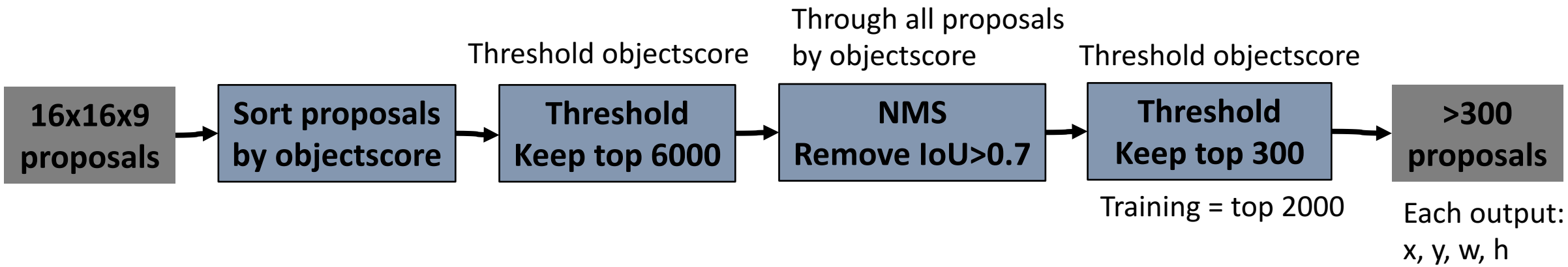
Anchor centers in image

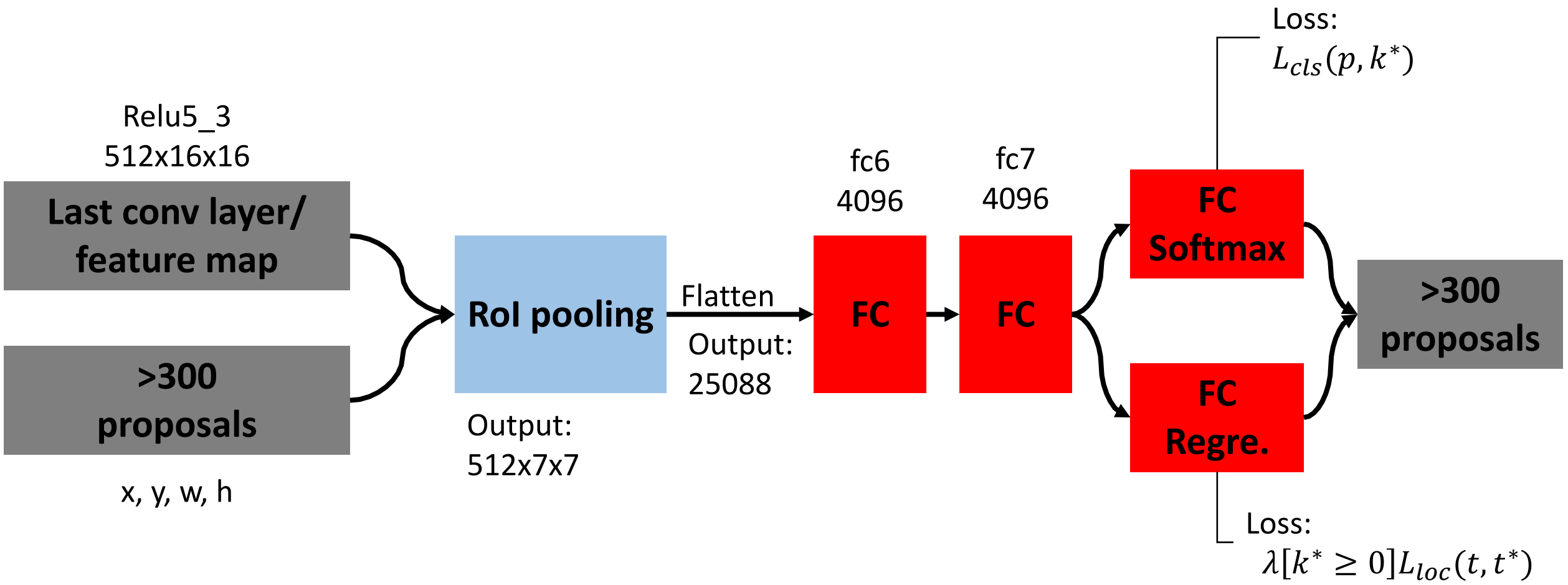


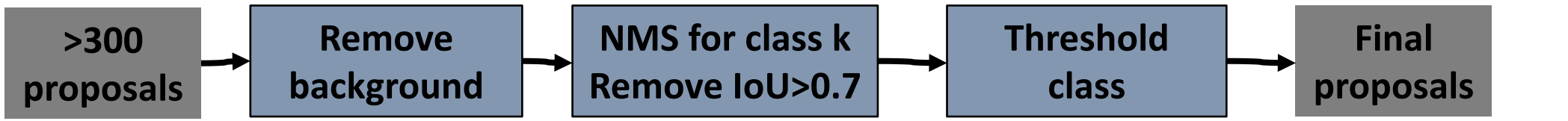
Each anchor=
3 scales
Ex. 10, 30, 50
3 aspect ratios
Ex 1:1, 1:2, 2:1

Anchors in image for 1 center



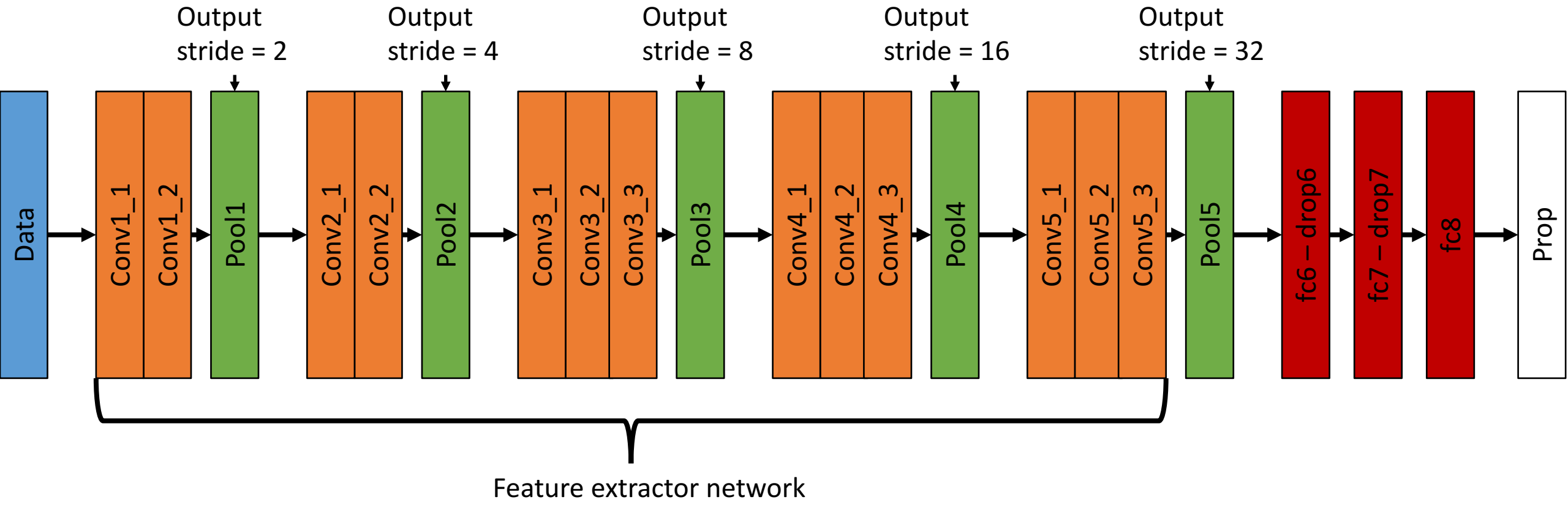

16x16x9 proposals
Sort proposals by objectscore
Threshold objectscore
Threshold Keep top 6000
Through all proposals by objectscore
NMS Remove IoU>0.7
Threshold objectscore
Threshold Keep top 300
Training = top 2000
>300 proposals
Each output: x, y, w, h

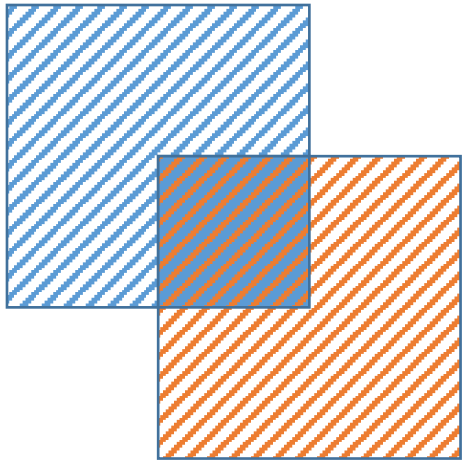

Relu5_3
512x16x16
Last conv layer/ feature map
>300 proposals
x, y, w, h
RoI pooling
Output:
512x7x7
Flatten
Output:
25088
fc6
4096
FC
fc7
4096
FC
FC Softmax
FC Regre.
Loss:
$L_{cls}(p, k^*)$
Loss:
$\lambda[k^* \geq 0]L_{loc}(t, t^*)$
>300 proposals



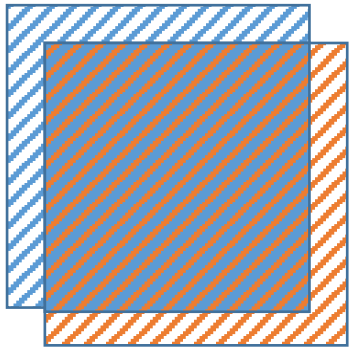
Not added in Faster R-CNN

x_{min}, x_{max}, y_{min}, y_{max}, Class, prob.

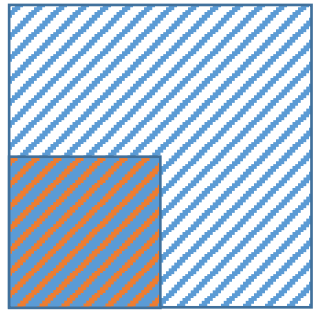

Output stride = 2
Output stride = 4
Output stride = 8
Output stride = 16
Output stride = 32
Data
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
fc6 – drop6
fc7 – drop7
fc8
Prop
Feature extractor network



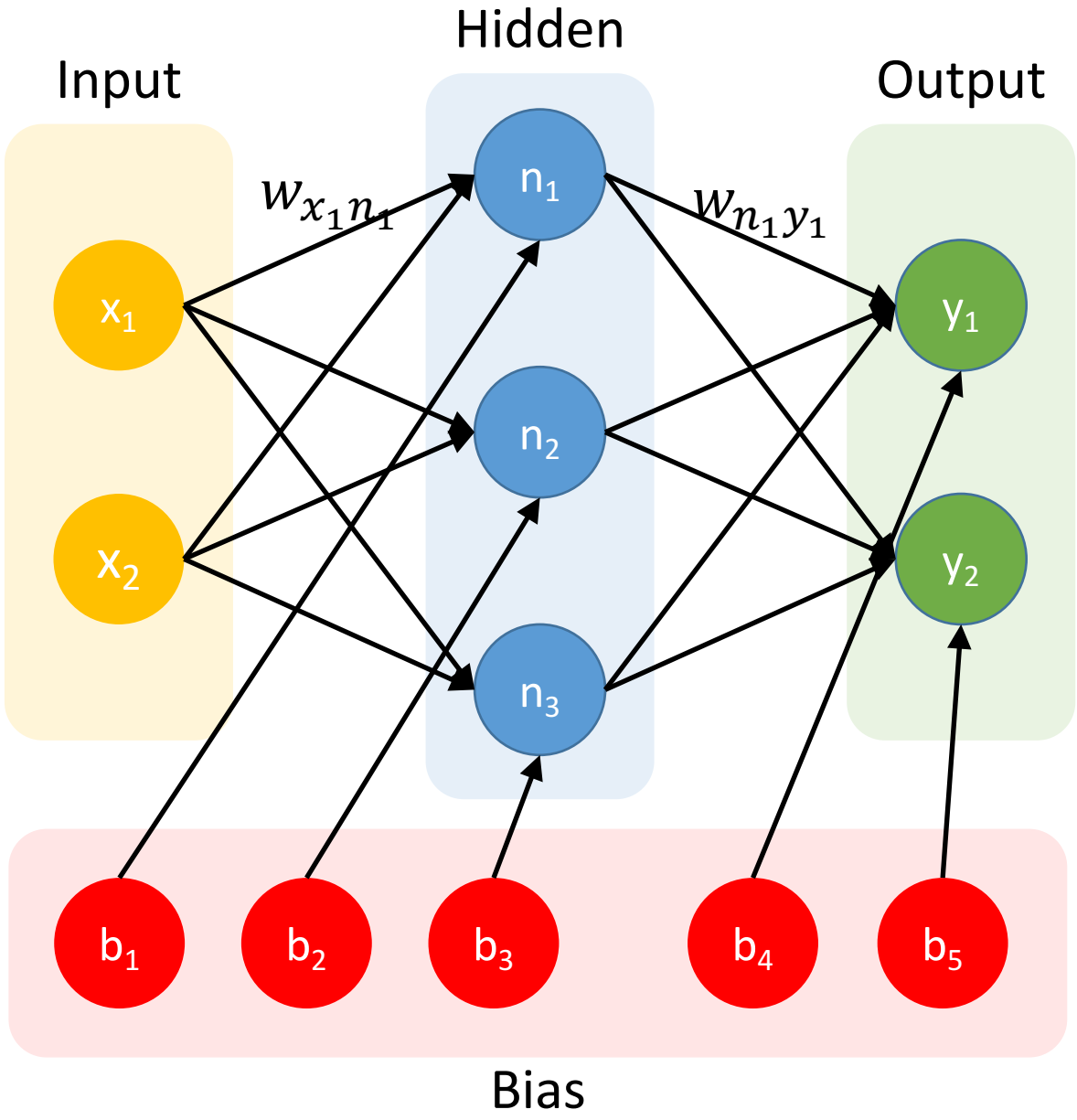
IoU = 0.1429

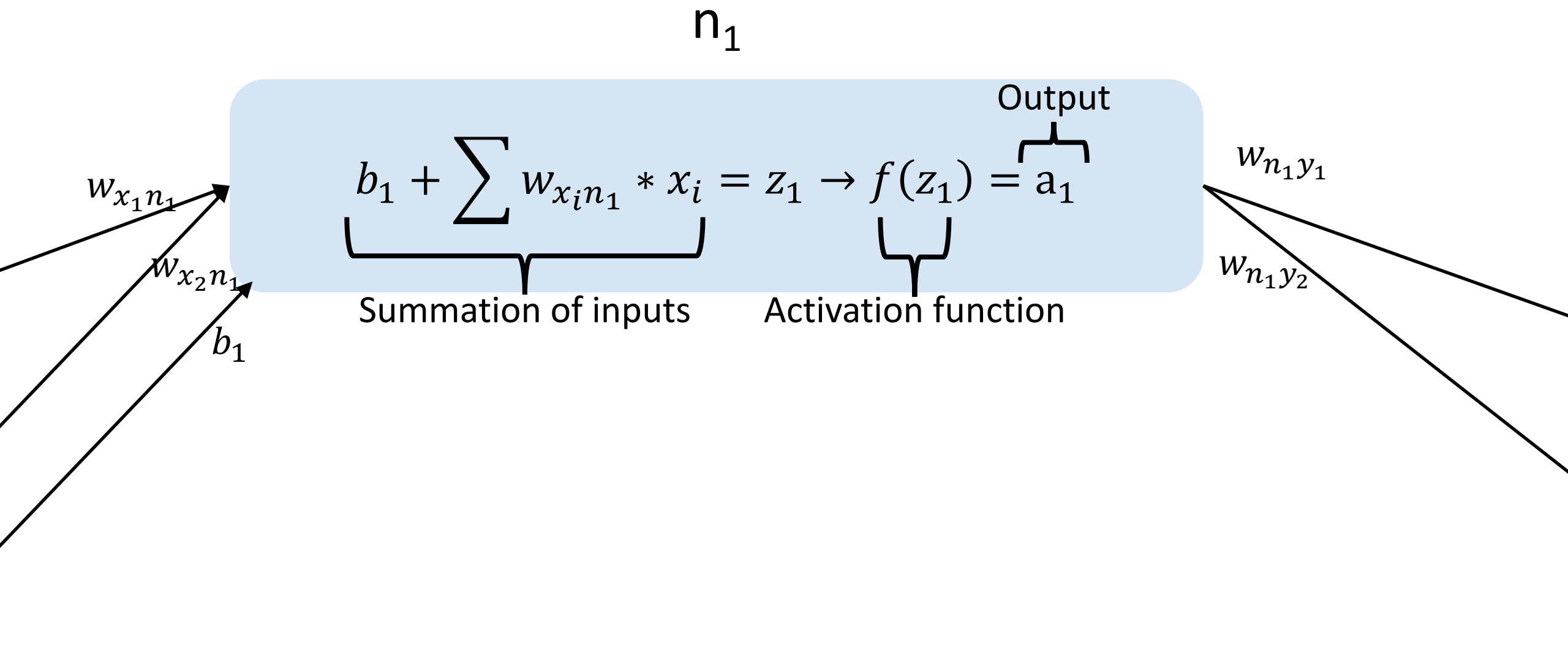


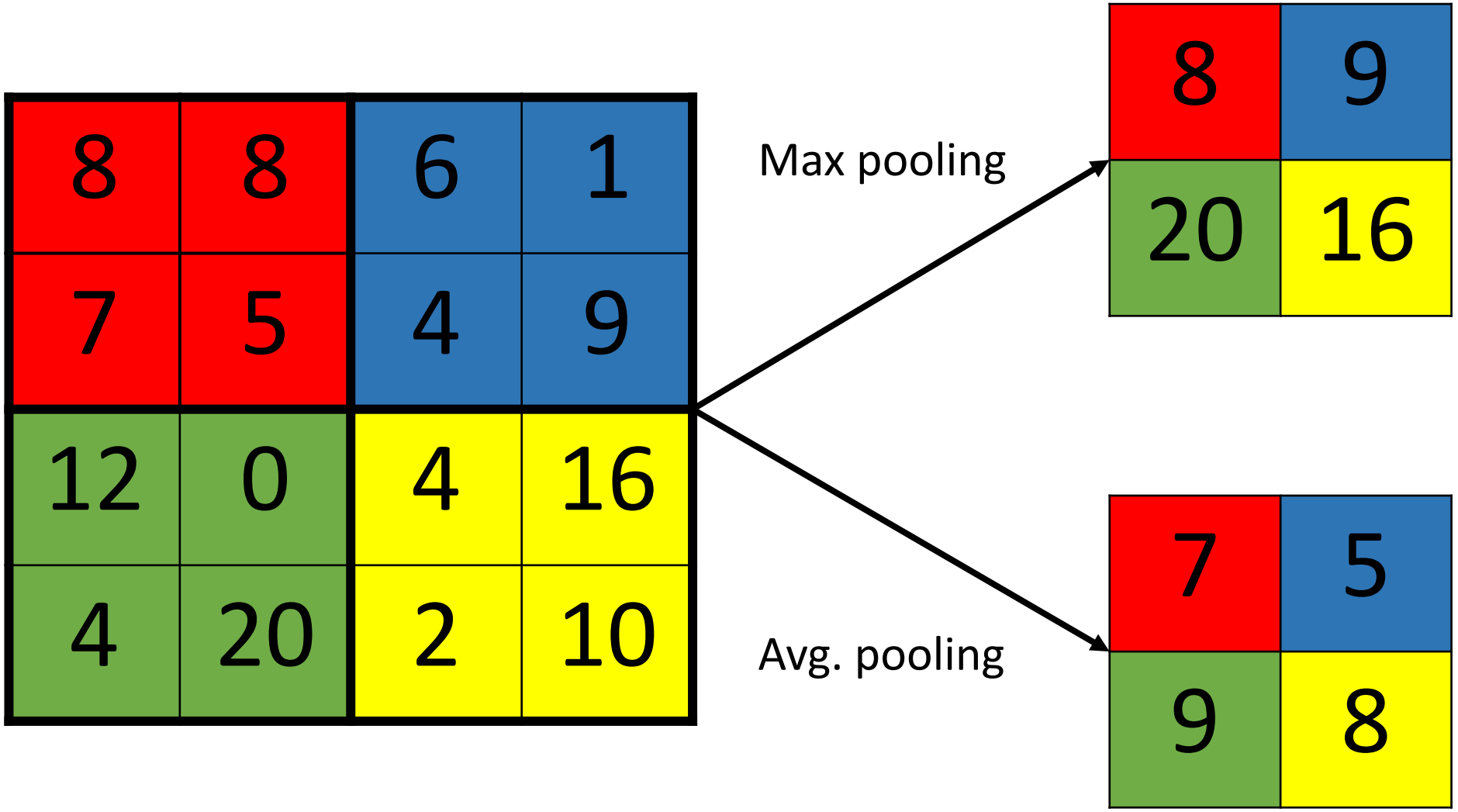
IoU = 0.6203

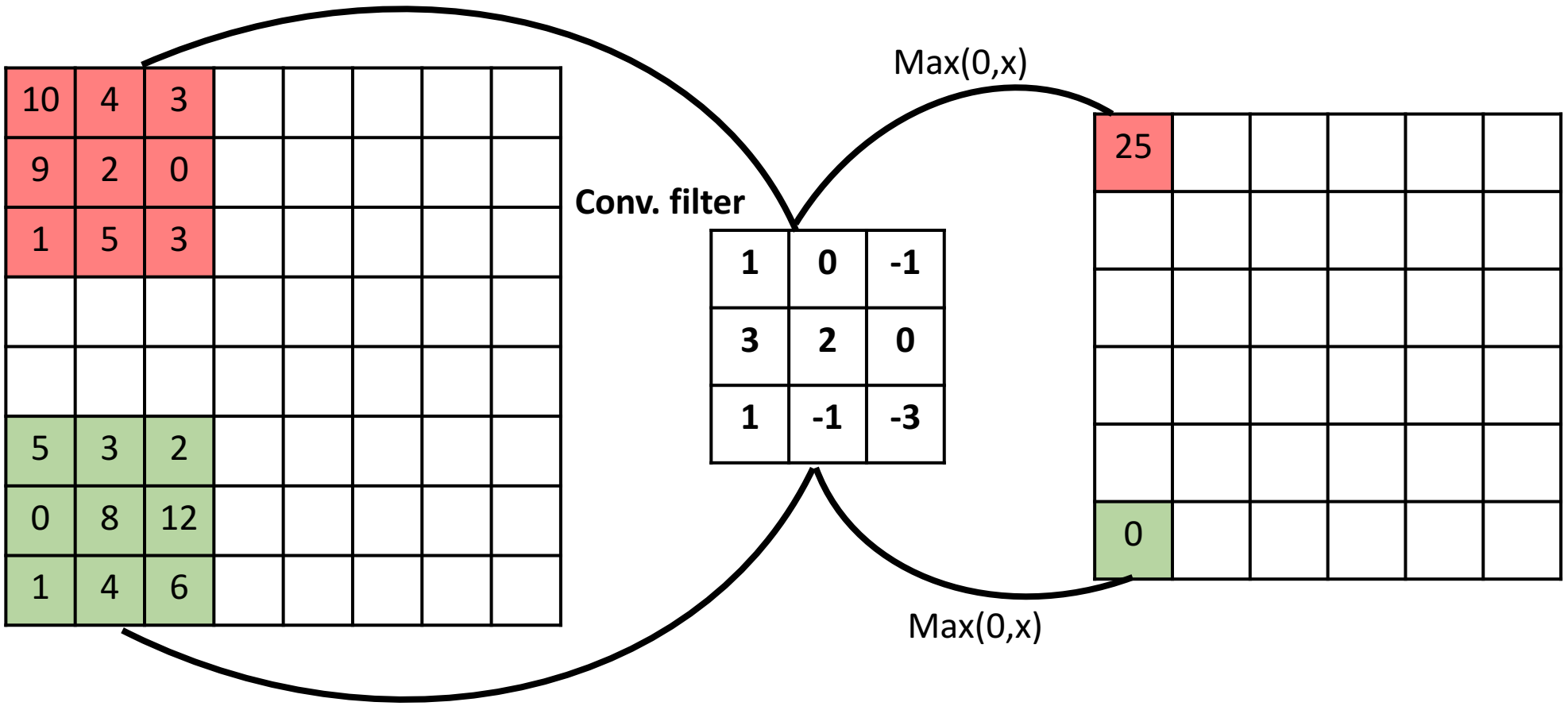


IoU = 0.2500


Hidden
Input
Output
$w_{x_1 n_1}$
$w_{n_1 y_1}$
x_1
x_2
n_1
n_2
n_3
y_1
y_2
b_1
b_2
b_3
b_4
b_5
Bias


n_1
$b_1 + \sum w_{x_i n_1} * x_i = z_1 \rightarrow f(z_1) = a_1$
Output
Summation of inputs
Activation function
$w_{x_1 n_1}$
$w_{x_2 n_1}$
b_1
$w_{n_1 y_1}$
$w_{n_1 y_2}$


8
8
6
1
7
5
4
9
12
0
4
16
4
20
2
10
Max pooling
Avg. pooling
8
9
20
16
7
5
9
8


10 4 3
9 2 0
1 5 3
5 3 2
0 8 12
1 4 6
Conv. filter
1 0 -1
3 2 0
1 -1 -3
Max(0,x)
25
0
Max(0,x)