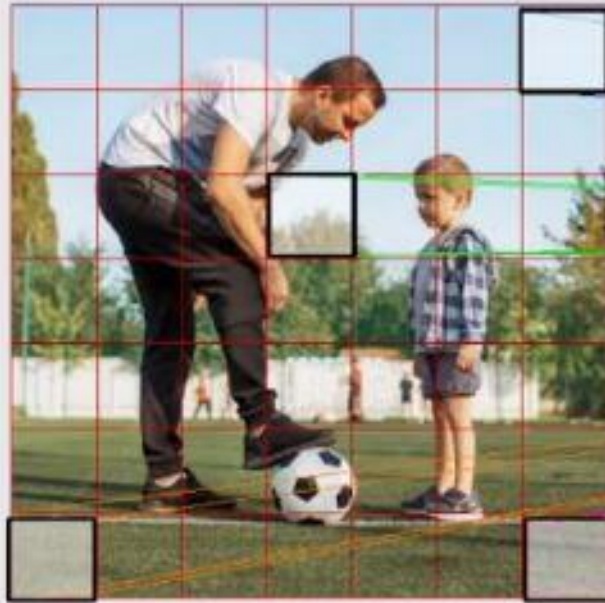


ROI Pooling and ROI Align Pooling

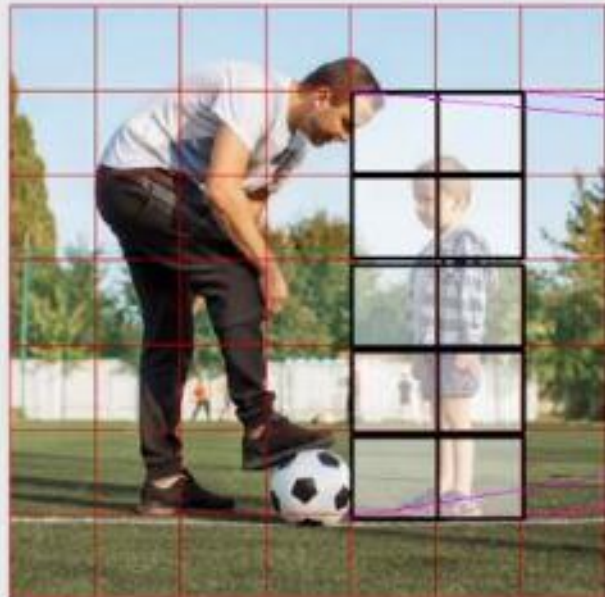


224×224  $7 \times 7 \times 512$
Feature map

7

7

After 5 Pooling Layers
 224×224 is represented by 7×7 values
(each value represents 32×32 pixels)

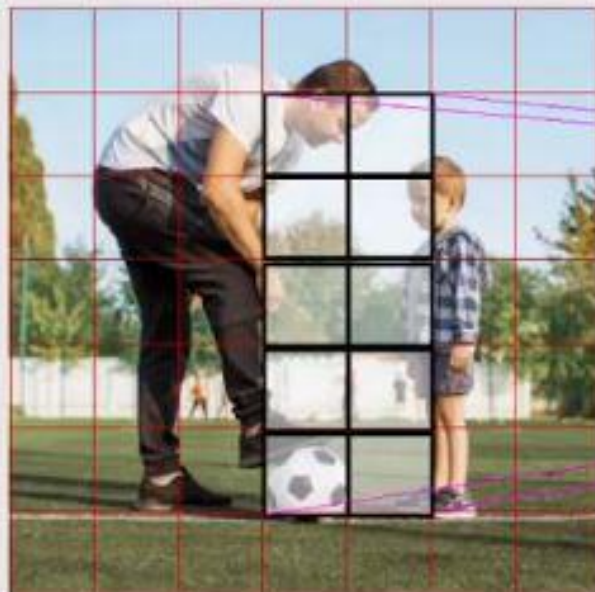
 224×224  $7 \times 7 \times 512$
Feature map

7

7

10 values in Feature Map are
representing the Boy

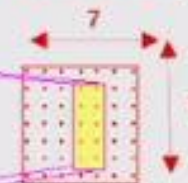
224 x 224



Effect of Shift and Crop of Feature Map

96

7x7x512
Feature map



Effect of shift of values in feature map.

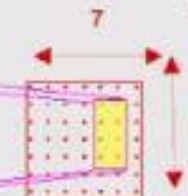
Missing Portion of the Boy

Adding portion of the father

[Error due to Quantization of Origin coordinates]



7x7x512
Feature map

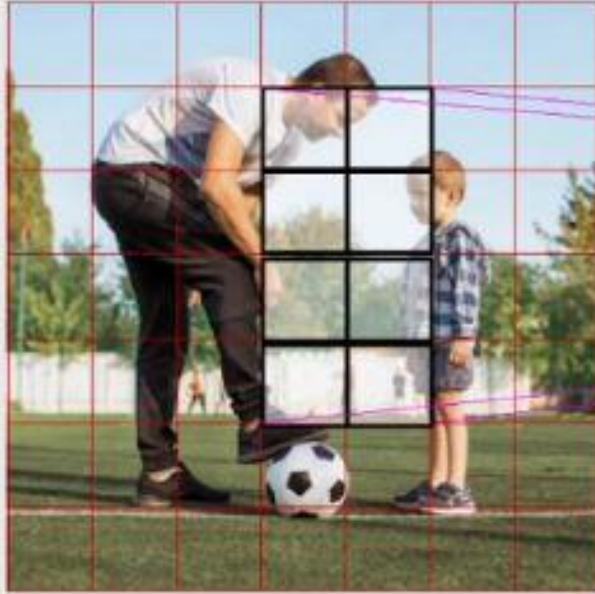


Effect of cropping some values in feature map.

Missing Portion of the Boy

[Error due to Quantization of Width and Height]

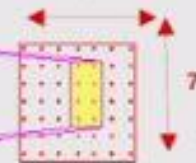
224 x 224



Effect of Shift and Crop of Feature Map

7x7x Chs
Feature map

7



Double Quantization Errors

[Error due to Quantization of Origin coordinates]

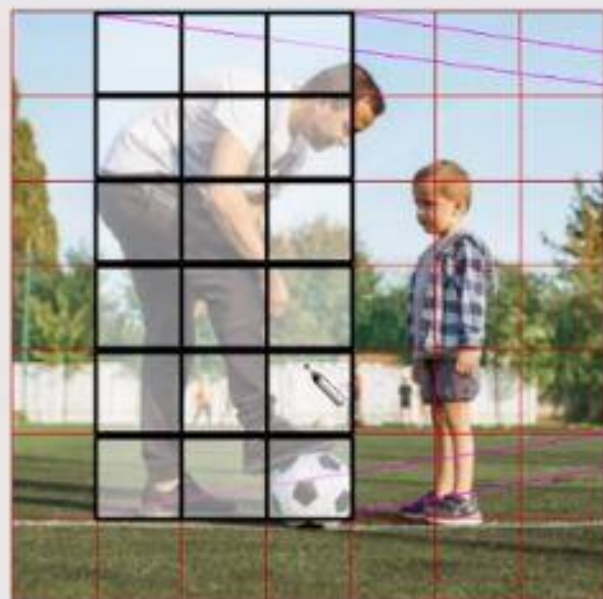
[Error due to Quantization of Width and Height]

Quantization is a source of ERROR

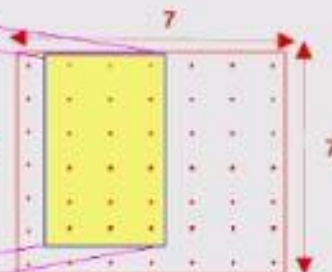
But

Why do we need to Quantize ?

Region Of Interest (ROI) Pooling to 2 x 2 size

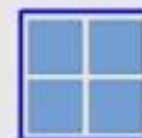
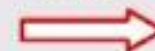


7x7xChs
Feature map

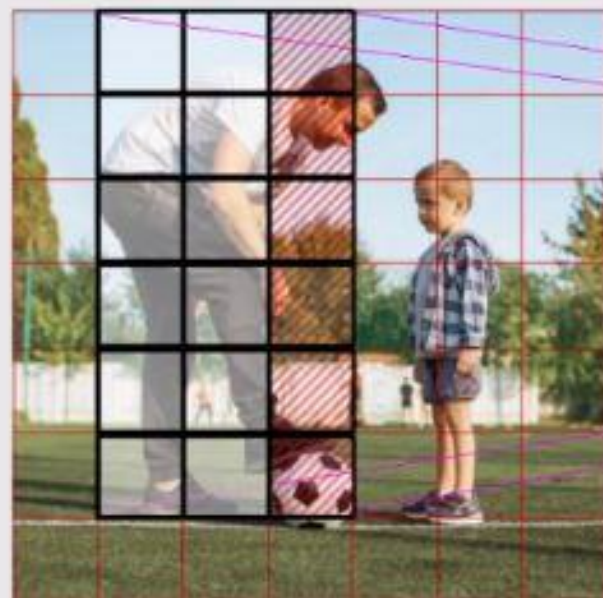


Object Feature Map
 $W=3, H=6$

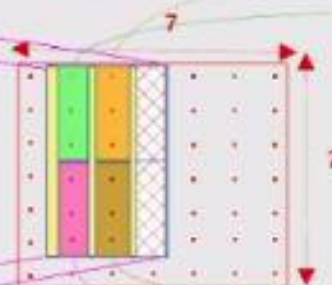
ROI Pooling
 2×2



Size after Pooling
 $W=2, H=3$

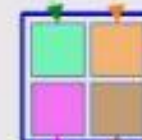


7x7xChs
Feature map



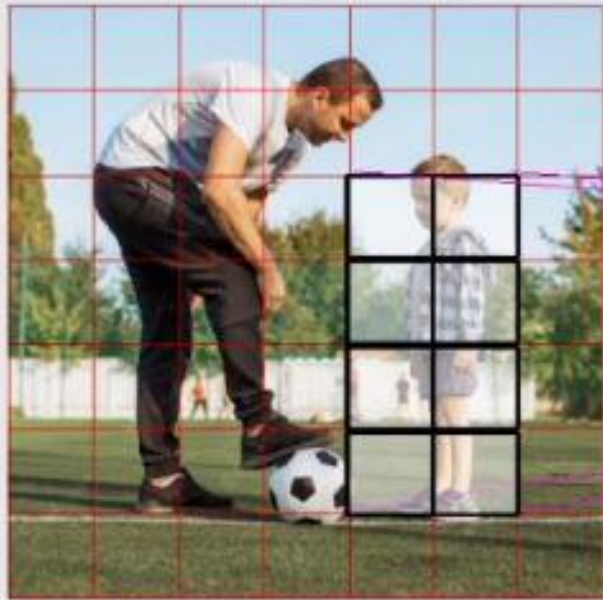
Object Feature Map
 $W=3, H=6$
 $3/2=1, 6/2=3$

ROI Pooling
 2×2

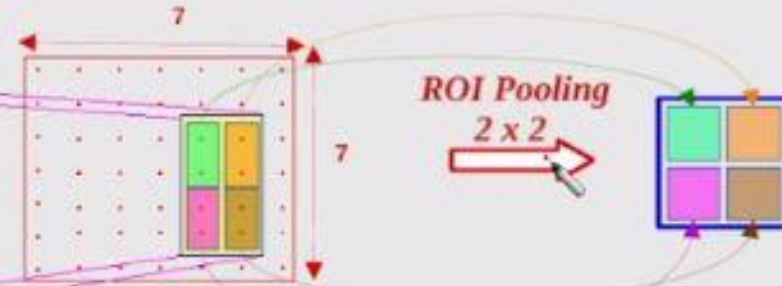


(Width Quantization error)

Region Of Interest (ROI) Pooling to 2 x 2 size



7x7xChs
Feature map

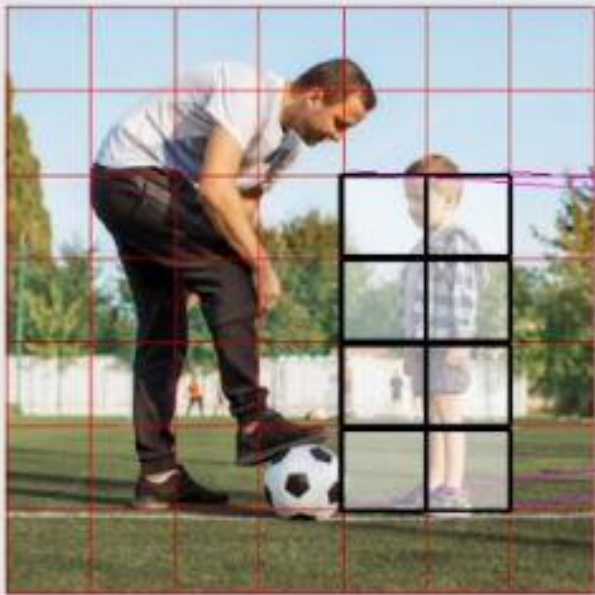


Object Feature Map
 $W=2, H=4$

Size after Pooling
 $W=2, H=3$

$2/2=1, 4/2=2$
(NO Quantization error BUT !!!)

Region Of Interest (ROI) Pooling to 2 x 2 size



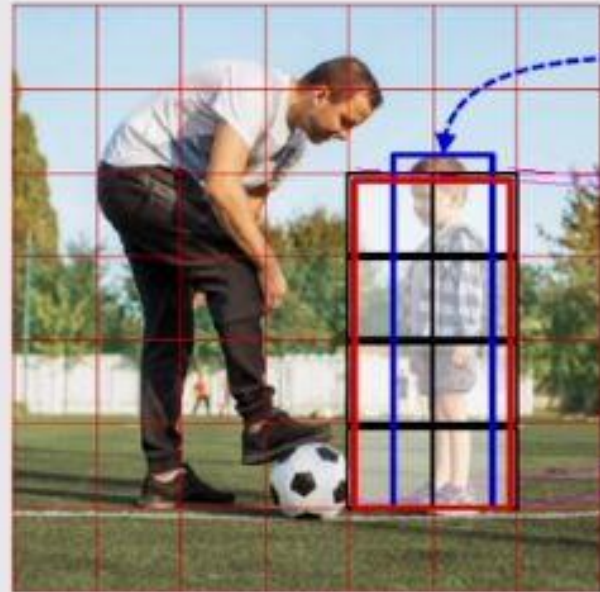
7x7xChs
Feature map



Object Feature Map
W=2, H=4

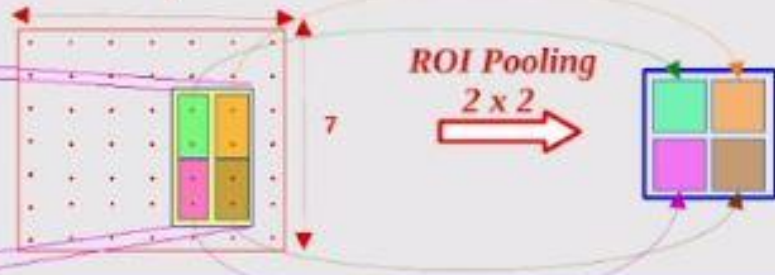
Size after Pooling
W=2, H=3

(NO Quantization error BUT !!!)



Ground Truth
Box

7x7xChs
Feature map



Box is NOT Properly aligned
Ground Truth Box (Blue)
Box Corresponding to selected values in Feature Map (Red)
Remember: Each value in Feature Map corresponds to 16 x 16 cell

ROI Pooling: Problem Definition

101

800 x 576 Image

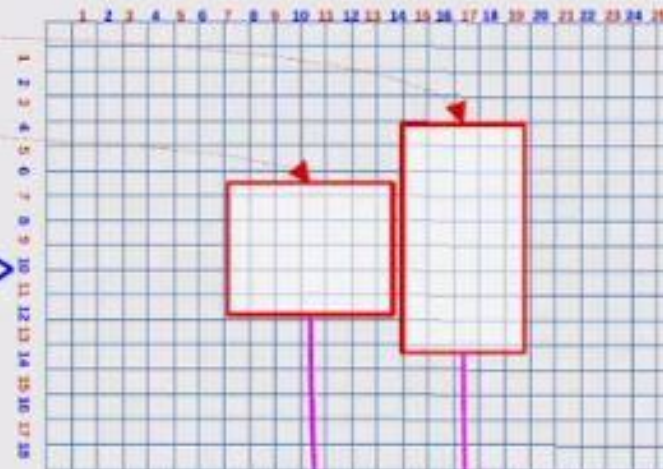


Region of Interest (ROI)

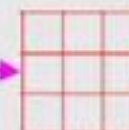
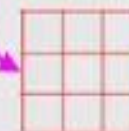
A Pre-Trained CNN Model
(with 5 Pooling Layers)

25 x 18 Feature Map

$$\frac{800}{32} \times \frac{576}{32}$$

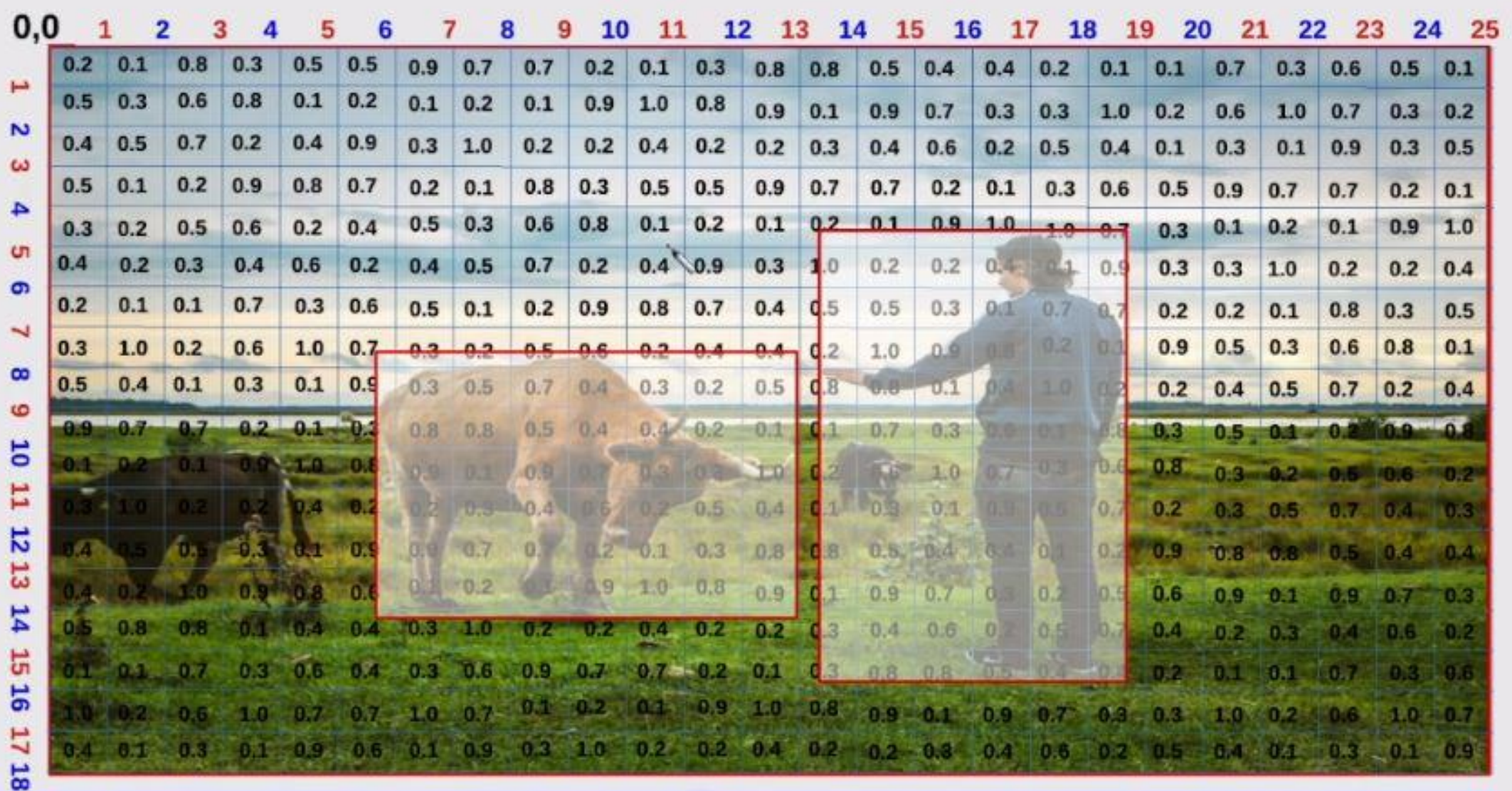


ROI Pooling
Output: 3x3
(irrespective of input size)



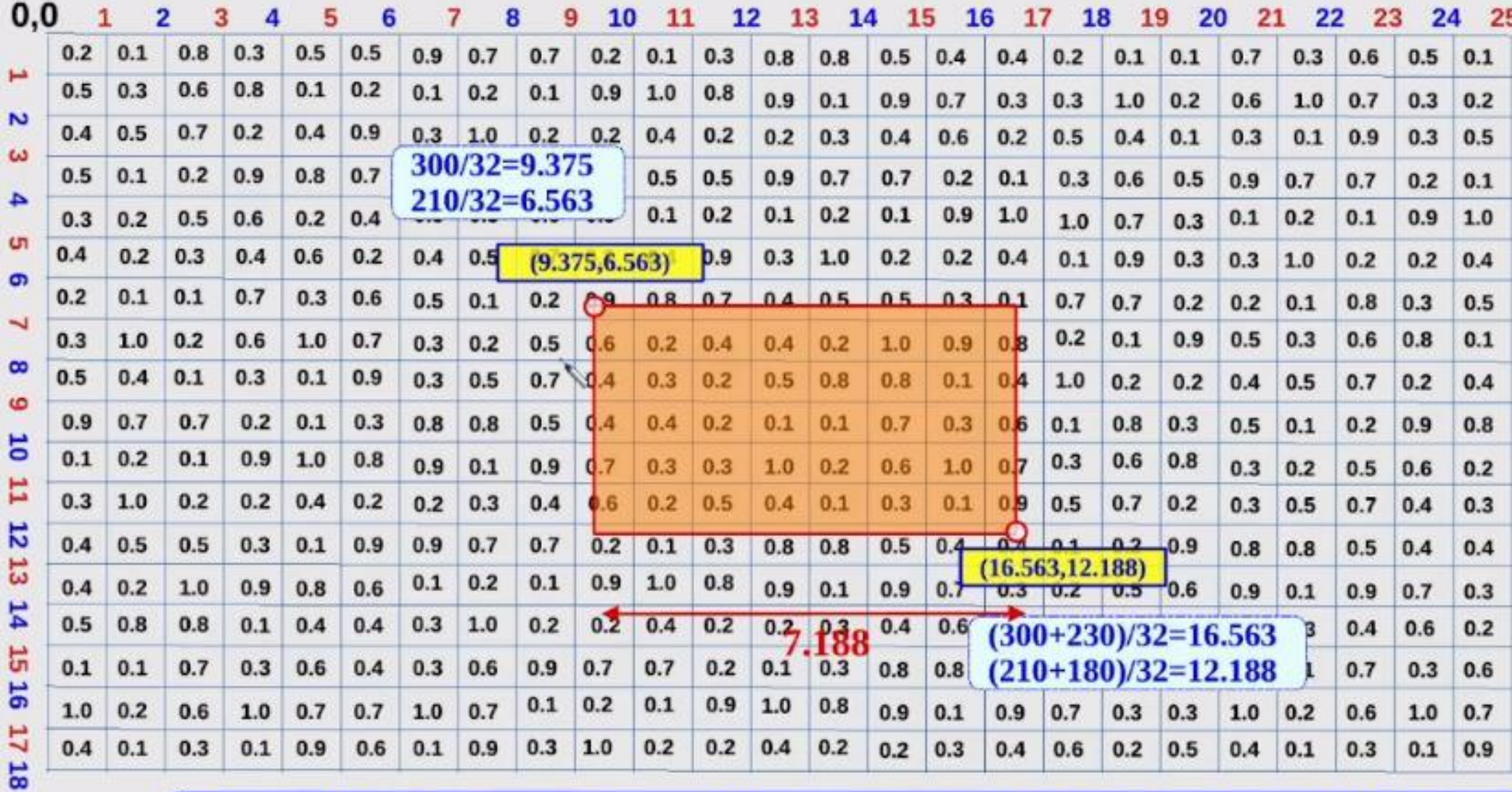
Fully Connected
NN Classifier

Feature Map (800/32 , 576/32) = 25 x 18



Feature Map of the whole image Width = 800, Height=576 , CNN with 5 Pooling Layers

Feature Map of selected object (Region of Interest [ROI])



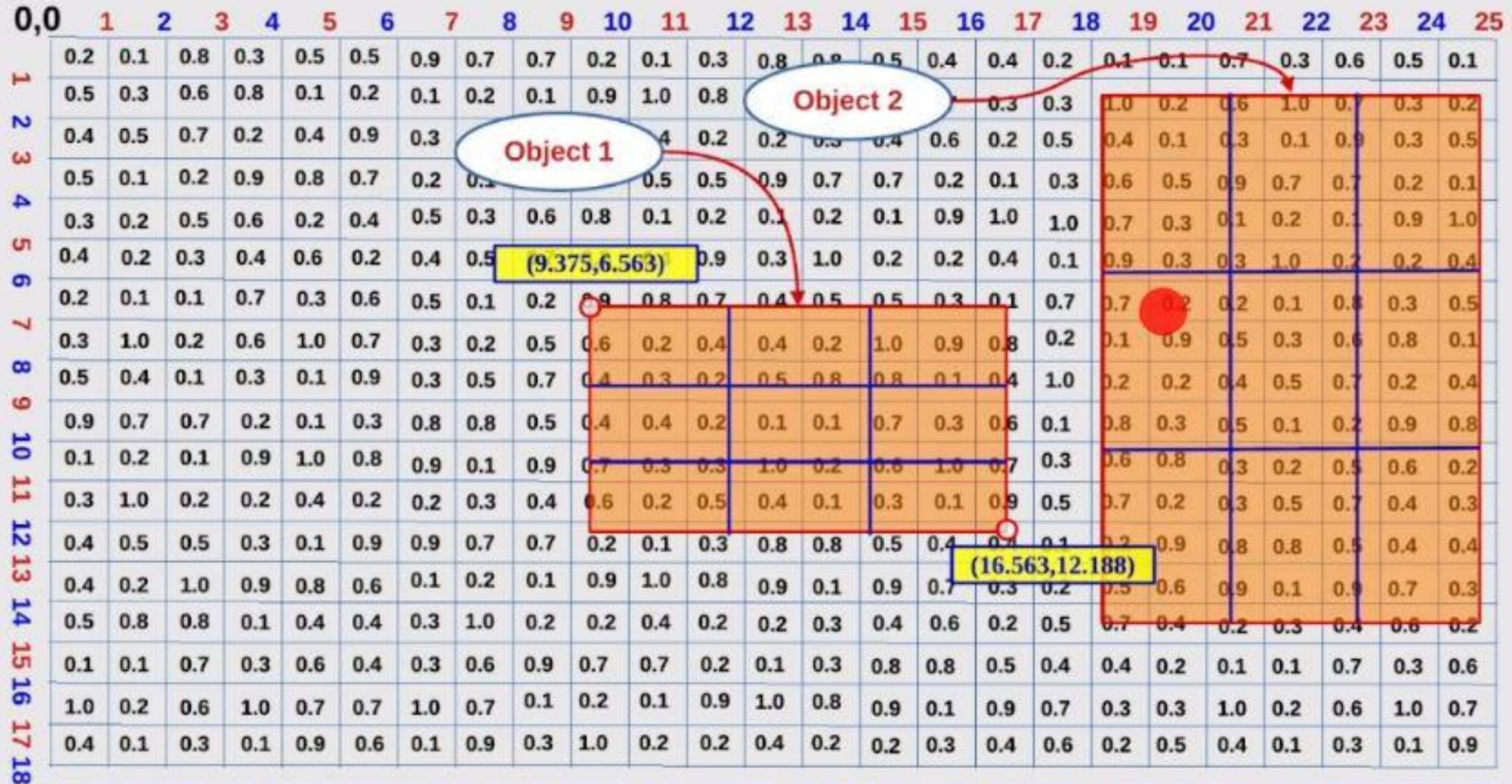
Feature Map Corresponding to Object at $x,y=(300,210)$ $w,h=(230,180)$

ROI Pooling Output is 3x3

0,0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1	0.2	0.1	0.8	0.3	0.5	0.5	0.9	0.7	0.7	0.2	0.1	0.3	0.8	0.8	0.5	0.4	0.4	0.2	0.1	0.1	0.7	0.3	0.6	0.5	0.1	
2	0.5	0.3	0.6	0.8	0.1	0.2	0.1	0.2	0.1	0.9	1.0	0.8	0.9	0.1	0.9	0.7	0.3	0.3	1.0	0.2	0.6	1.0	0.7	0.3	0.2	
3	0.4	0.5	0.7	0.2	0.4	0.9	0.3	1.0	0.2	0.2	0.4	0.2	0.2	0.3	0.4	0.6	0.2	0.5	0.4	0.1	0.3	0.1	0.9	0.3	0.5	
4	0.5	0.1	0.2	0.9	0.8	0.7	0.2	0.1	0.8	0.3	0.5	0.5	0.9	0.7	0.7	0.2	0.1	0.3	0.6	0.5	0.9	0.7	0.7	0.2	0.1	
5	0.3	0.2	0.5	0.6	0.2	0.4	0.5	0.3	0.6	0.8	0.1	0.2	0.1	0.2	0.1	0.9	1.0	1.0	0.7	0.3	0.1	0.2	0.1	0.9	1.0	
6	0.4	0.2	0.3	0.4	0.6	0.2	0.4	0.5	(9.375,6.563)				0.9	0.3	1.0	0.2	0.2	0.4	0.1	0.9	0.3	0.3	1.0	0.2	0.2	0.4
7	0.2	0.1	0.1	0.7	0.3	0.6	0.5	0.1	0.2	0.9	0.8	0.7	0.4	0.5	0.5	0.3	0.1	0.7	0.7	0.2	0.2	0.1	0.8	0.3	0.5	
8	0.3	1.0	0.2	0.6	1.0	0.7	0.3	0.2	0.5	0.6	0.2	0.4	0.4	0.2	1.0	0.9	0.8	0.2	0.1	0.9	0.5	0.3	0.6	0.8	0.1	
9	0.5	0.4	0.1	0.3	0.1	0.9	0.3	0.5	0.7	0.4	0.3	0.2	0.5	0.8	0.8	0.1	0.4	1.0	0.2	0.2	0.4	0.5	0.7	0.2	0.4	
10	0.9	0.7	0.7	0.2	0.1	0.3	0.8	0.8	0.5	0.4	0.4	0.2	0.1	0.1	0.7	0.3	0.6	0.1	0.8	0.3	0.5	0.1	0.2	0.9	0.8	
11	0.1	0.2	0.1	0.9	1.0	0.8	0.9	0.1	0.9	0.7	0.3	0.3	1.0	0.2	0.8	1.0	0.7	0.3	0.6	0.8	0.3	0.2	0.5	0.6	0.2	
12	0.3	1.0	0.2	0.2	0.4	0.2	0.2	0.3	0.4	0.6	0.2	0.5	0.4	0.1	0.3	0.1	0.9	0.5	0.7	0.2	0.3	0.5	0.7	0.4	0.3	
13	0.4	0.5	0.5	0.3	0.1	0.9	0.9	0.7	0.7	0.2	0.1	0.3	0.8	0.8	0.5	0.4	0.4	0.1	0.2	0.9	0.8	0.8	0.5	0.4	0.4	
14	0.4	0.2	1.0	0.9	0.8	0.6	0.1	0.2	0.1	0.9	1.0	0.8	0.9	0.1	0.9	0.7	0.3	0.2	0.5	0.6	0.9	0.1	0.9	0.7	0.3	
15	0.5	0.8	0.8	0.1	0.4	0.4	0.3	1.0	0.2	0.2	0.4	0.2	0.2	0.3	0.4	0.6	0.2	0.5	0.7	0.4	0.2	0.3	0.4	0.6	0.2	
16	0.1	0.1	0.7	0.3	0.6	0.4	0.3	0.6	0.9	0.7	0.7	0.2	0.1	0.3	0.8	0.8	0.5	0.4	0.4	0.2	0.1	0.1	0.7	0.3	0.6	
17	1.0	0.2	0.6	1.0	0.7	0.7	1.0	0.7	0.1	0.2	0.1	0.9	1.0	0.8	0.9	0.1	0.9	0.7	0.3	0.3	1.0	0.2	0.6	1.0	0.7	
18	0.4	0.1	0.3	0.1	0.9	0.6	0.1	0.9	0.3	1.0	0.2	0.2	0.4	0.2	0.2	0.3	0.4	0.6	0.2	0.5	0.4	0.1	0.3	0.1	0.9	

Feature Map of ROI (object) will be pooled to 3x3 (Align Pooling)

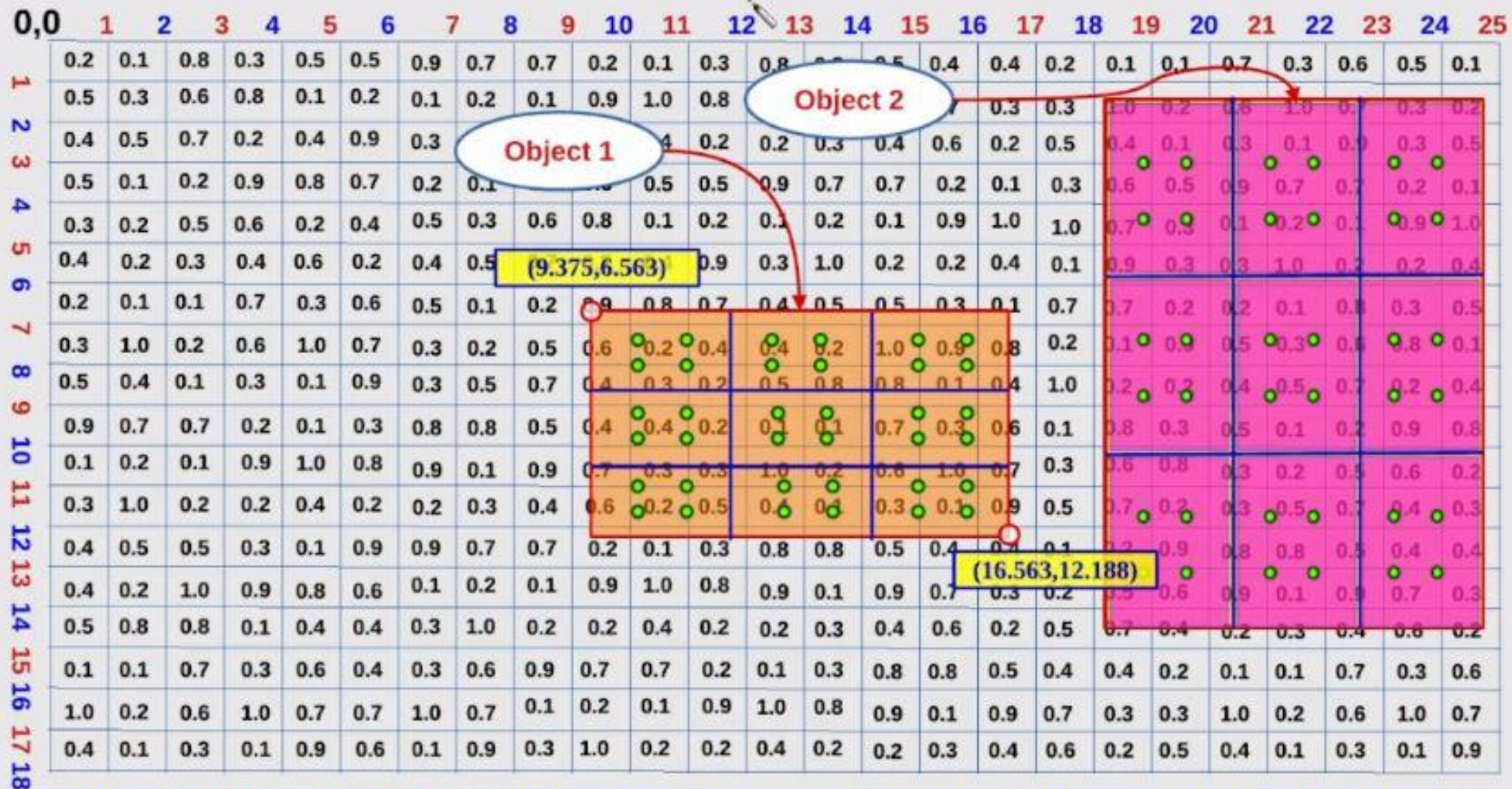
ROI Pooling Output is 3x3 (Irrespective of Object Size / Aspect Ratio)



Feature Map Corresponding to two Objects with Different Sizes (all will be Pooled to 3x3)

Sampling Points "S_x" represent values inside each cell (4 Points)

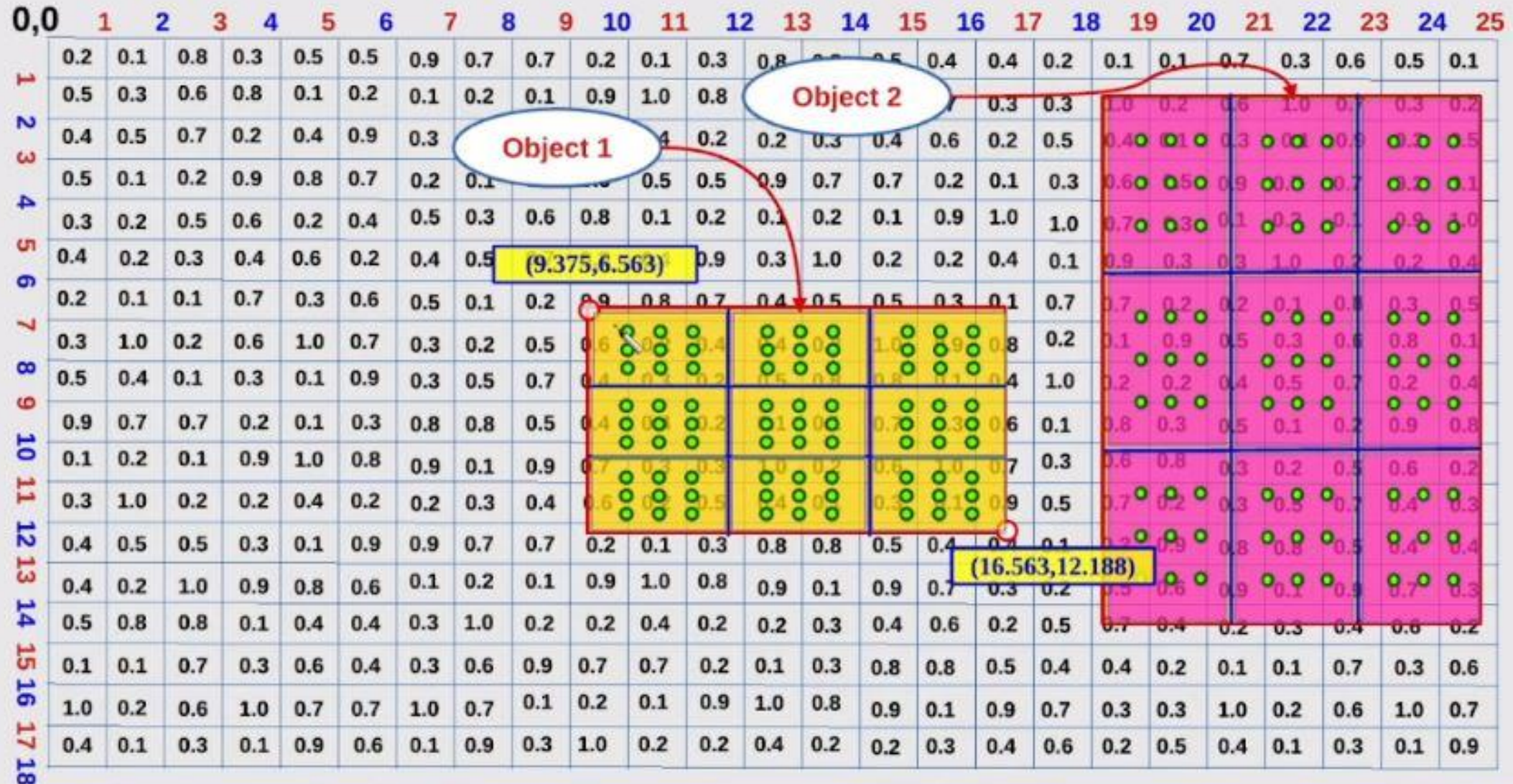
107



Select 2x2 Sampling Points (Calculate Corresponding Values then apply Max Pooling)

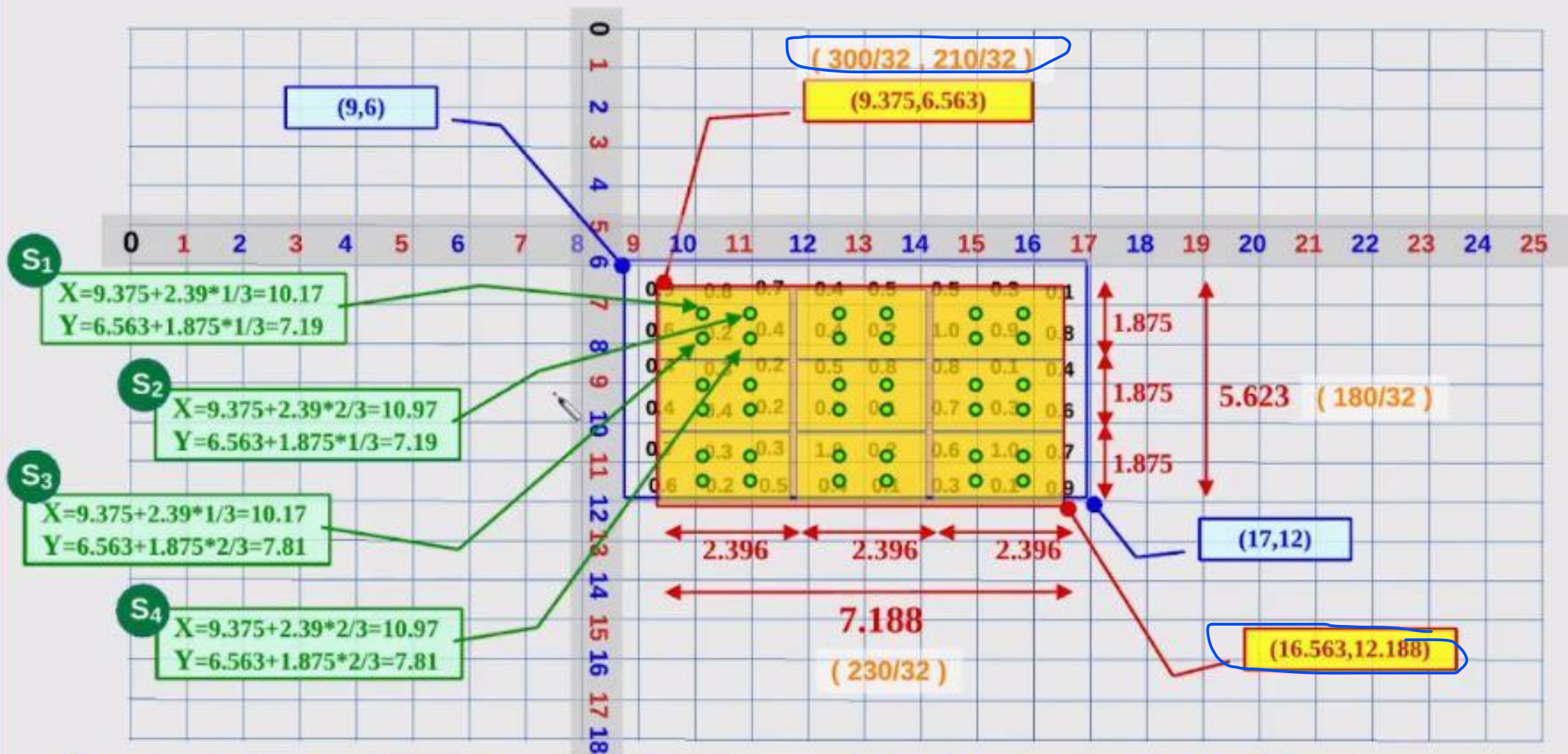
Sampling Points " S_x " represent values inside each cell (9 Points)

108



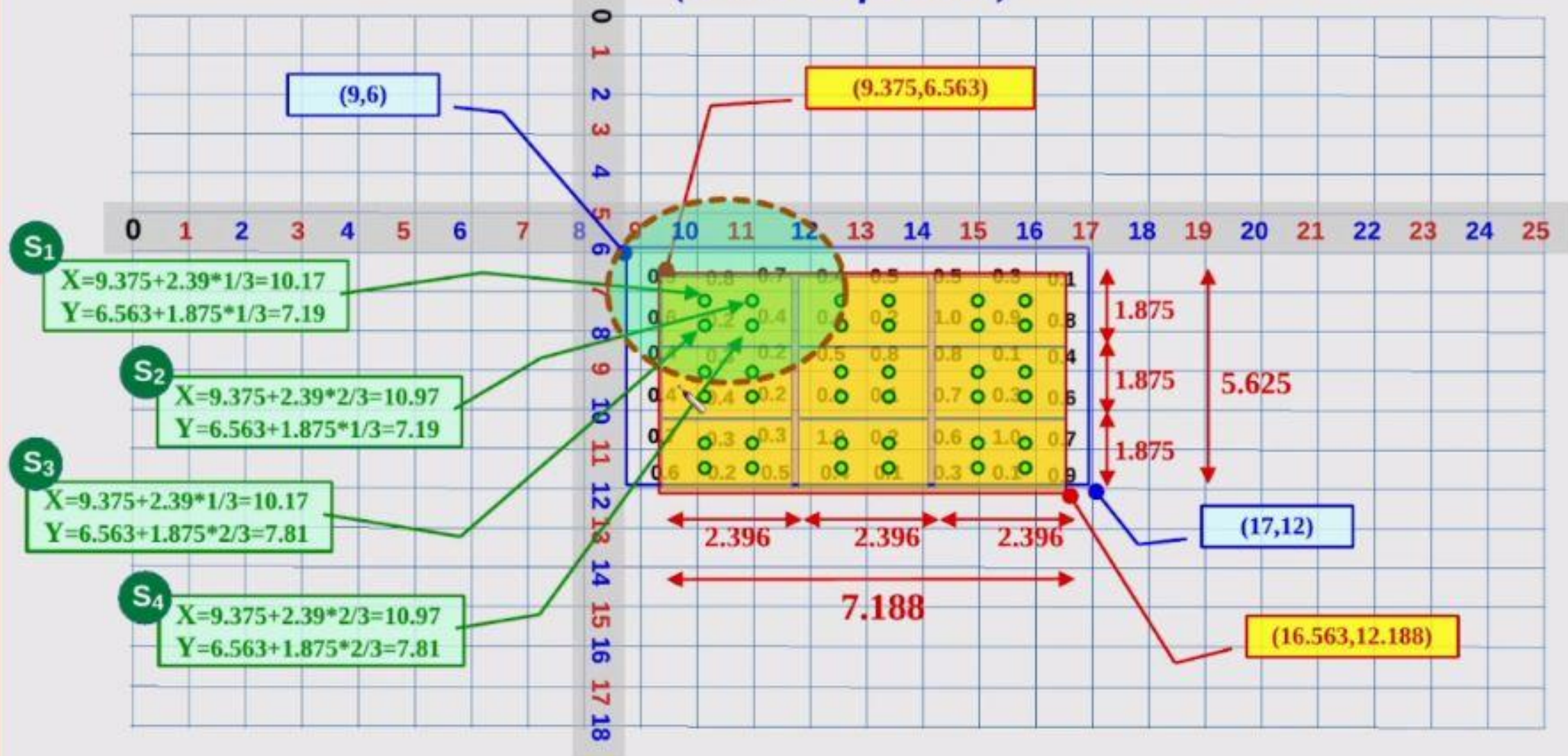
Select 3x3 Sampling Points (Calculate Corresponding Values then apply Max Pooling)

Sampling Points Coordinates



Feature Map Corresponding to Object at $x,y = (300,210)$ $w,h = (230,180)$, 5 Times Pooling

Sampling Points Coordinates (Focus on first cell)

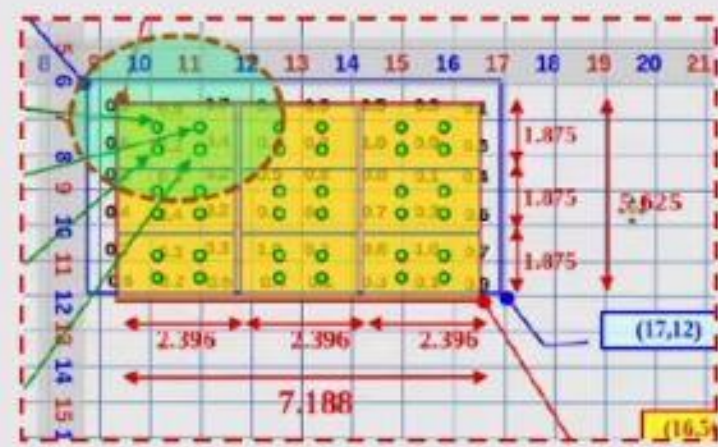
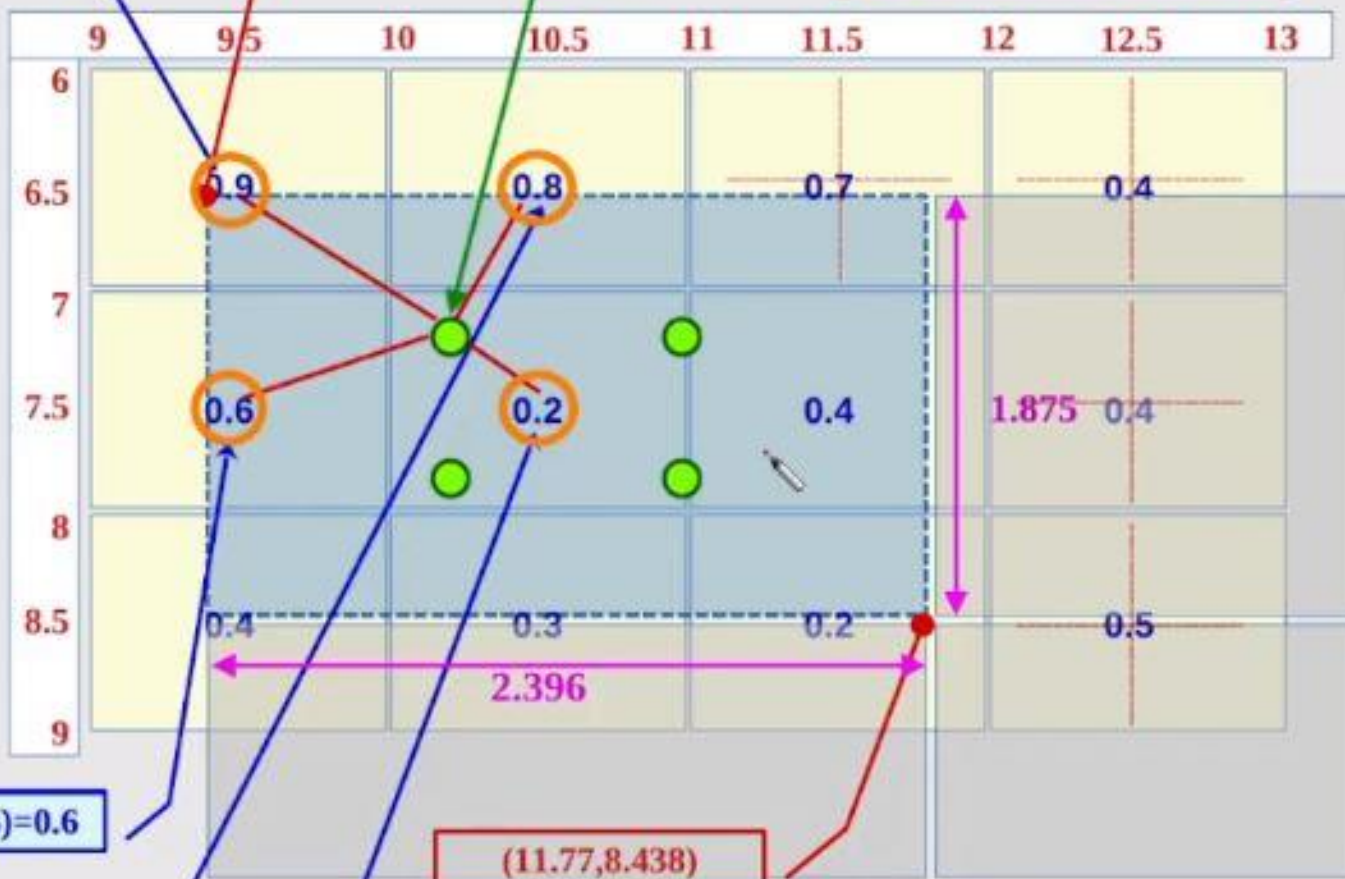


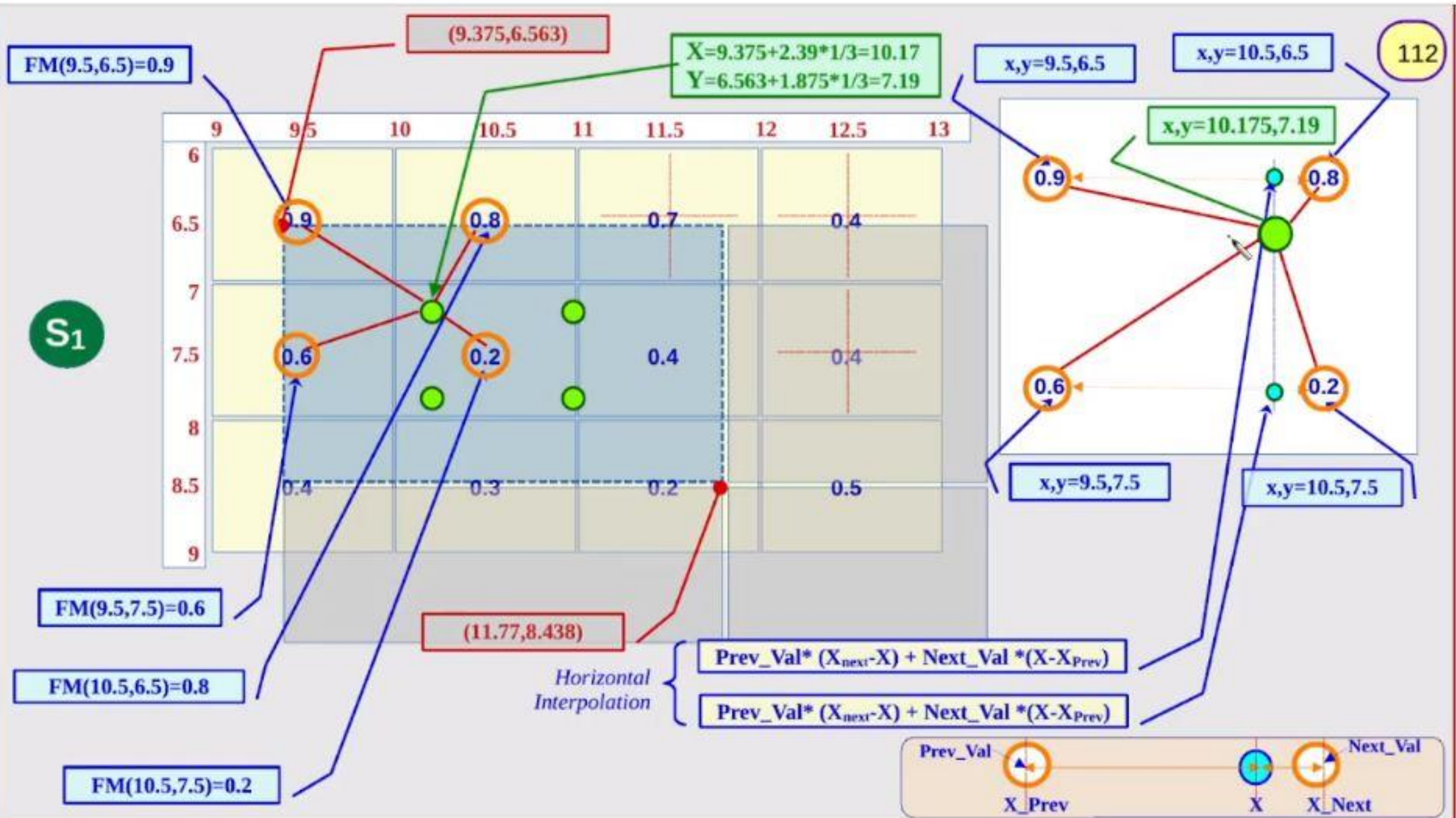
FM(9.5,6.5)=0.9

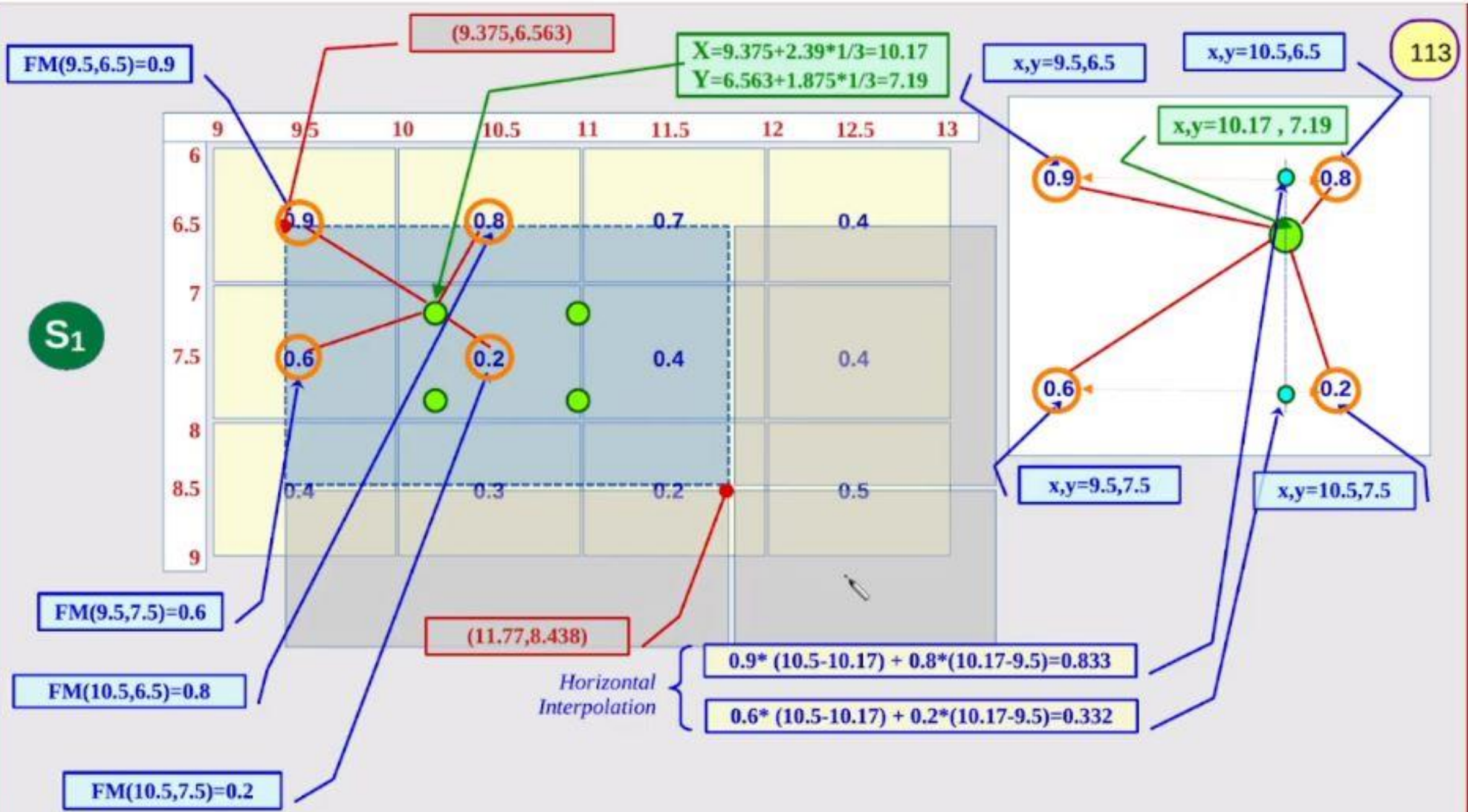
(9.375,6.563)

$X=9.375+2.39*1/3=10.17$
 $Y=6.563+1.875*1/3=7.19$

S₁







$$FM(10.5,6.5)=0.8$$

(9.375,6.563)

$$X=9.375+2.39*2/3=10.97$$

$$Y=6.563+1.875*1/3=7.19$$

x,y=10.5,6.5

x,y=11.5,6.5

115

x,y=10.97, 7.19

0.8

0.7

0.2

0.4

x,y=10.5,7.5

x,y=11.5,7.5

S₂

$$FM(10.5,7.5)=0.2$$

$$FM(11.5,6.5)=0.7$$

$$FM(11.5,7.5)=0.4$$

(11.77,8.438)

Horizontal
Interpolation

$$0.8 * (11.5 - 10.97) + 0.7 * (10.97 - 10.5) = 0.753$$

$$0.2 * (11.5 - 10.97) + 0.4 * (10.97 - 10.5) = 0.294$$

Vertical
Interpolation

$$0.753 * (7.5 - 7.19) + 0.294 * (7.19 - 6.5) = 0.436$$

$$FM(9.5,7.5)=0.6$$

$$(9.375,6.563)$$

$$X=9.375+2.39*1/3=10.17$$

$$Y=6.563+1.875*2/3=7.81$$

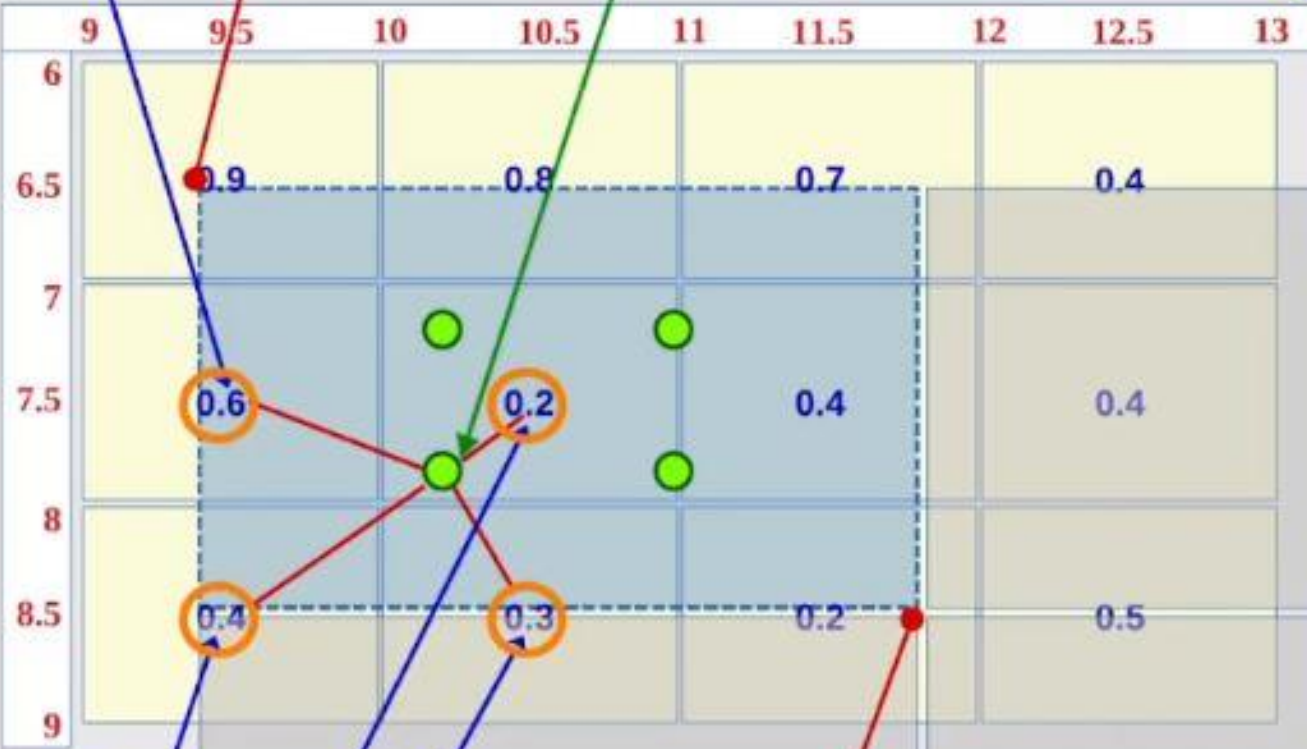
$$x,y=9.5,7.5$$

$$x,y=10.5,7.5$$

$$116$$

$$x,y=10.17, 7.81$$

S_3



$$0.6$$

$$0.2$$

$$0.4$$

$$0.3$$

$$x,y=9.5,8.5$$

$$x,y=10.5,8.5$$

$$FM(9.5,8.5)=0.4$$

$$FM(10.5,7.5)=0.2$$

$$FM(10.5,8.5)=0.3$$

$$(11.77,8.438)$$

Horizontal
Interpolation

$$0.6 * (10.5-10.17) + 0.2 * (10.17-9.5)=0.332$$

$$0.4 * (10.5-10.17) + 0.3 * (10.17-9.5)=0.333$$

Vertical
Interpolation

$$0.332 * (8.5-7.81) + 0.333 * (7.81-7.5)=0.332$$

$$FM(10.5, 7.5) = 0.2$$

$$(9.375, 6.563)$$

$$X = 9.375 + 2.39 \cdot \frac{2}{3} = 10.97$$

$$Y = 6.563 + 1.875 \cdot \frac{2}{3} = 7.81$$

$$x, y = 10.5, 7.5$$

$$x, y = 11.5, 7.5$$

$$117$$

$$x, y = 10.97, 7.81$$

$$0.2$$

$$0.4$$

$$0.3$$

$$0.2$$

$$x, y = 10.5, 8.5$$

$$x, y = 11.5, 8.5$$

S₄

$$FM(10.5, 8.5) = 0.3$$

$$(11.77, 8.438)$$

$$FM(11.5, 7.5) = 0.4$$

$$FM(11.5, 8.5) = 0.2$$

Horizontal
Interpolation

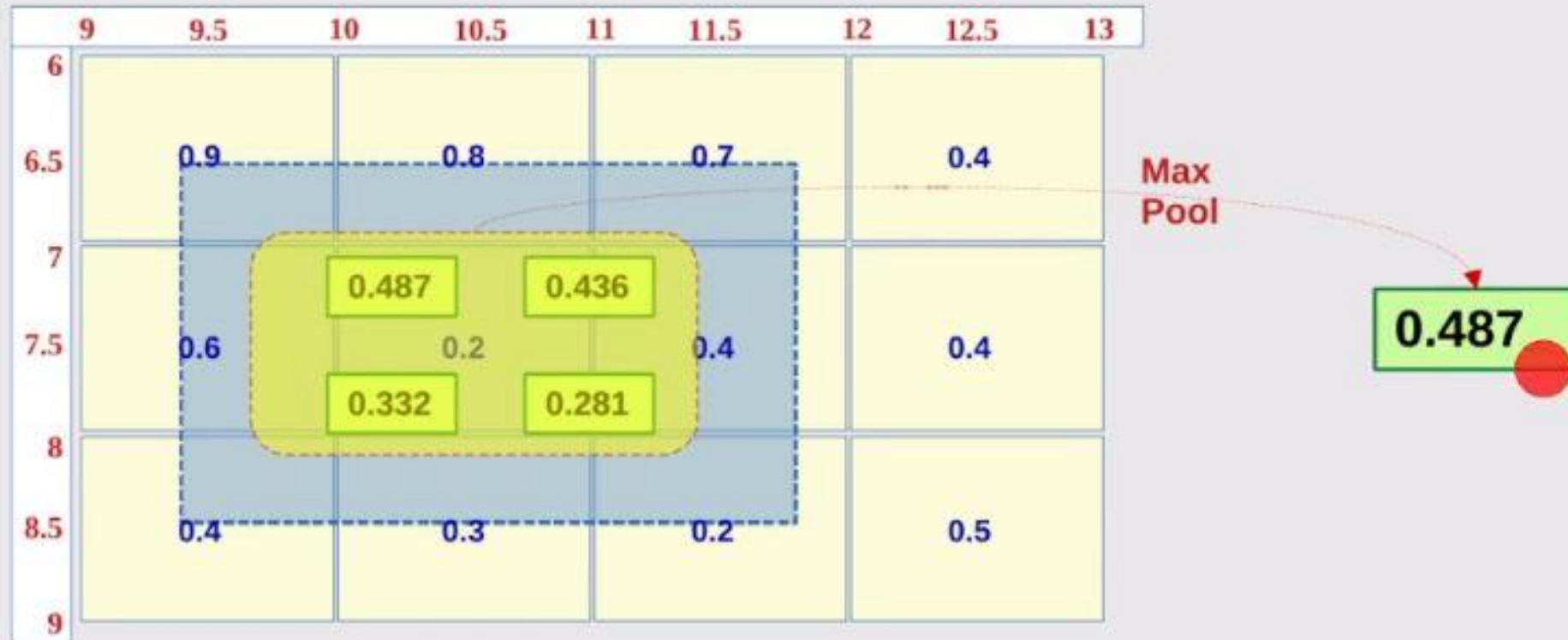
$$0.2 \cdot (11.5 - 10.97) + 0.4 \cdot (10.97 - 10.5) = 0.294$$

$$0.3 \cdot (11.5 - 10.97) + 0.2 \cdot (10.97 - 10.5) = 0.253$$

Vertical
Interpolation

$$0.294 \cdot (8.5 - 7.81) + 0.253 \cdot (7.81 - 7.5) = 0.281$$

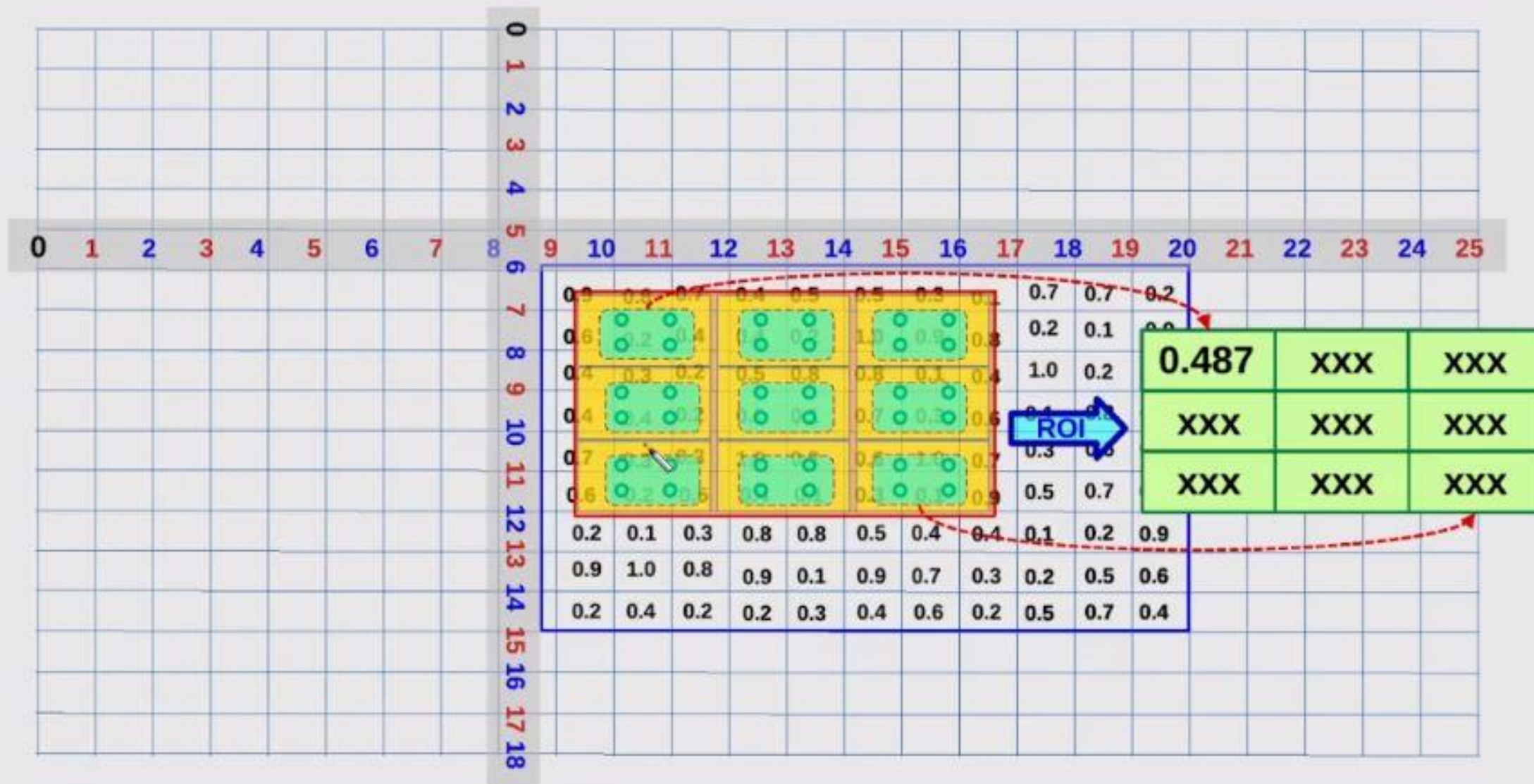
Apply Max-Pooling on the Calculated Sampling Points



Repeat for All other Cells Within the Region Of Interest (ROI) to obtain 3x3 values

3x3 ROI Align for selected Object

119



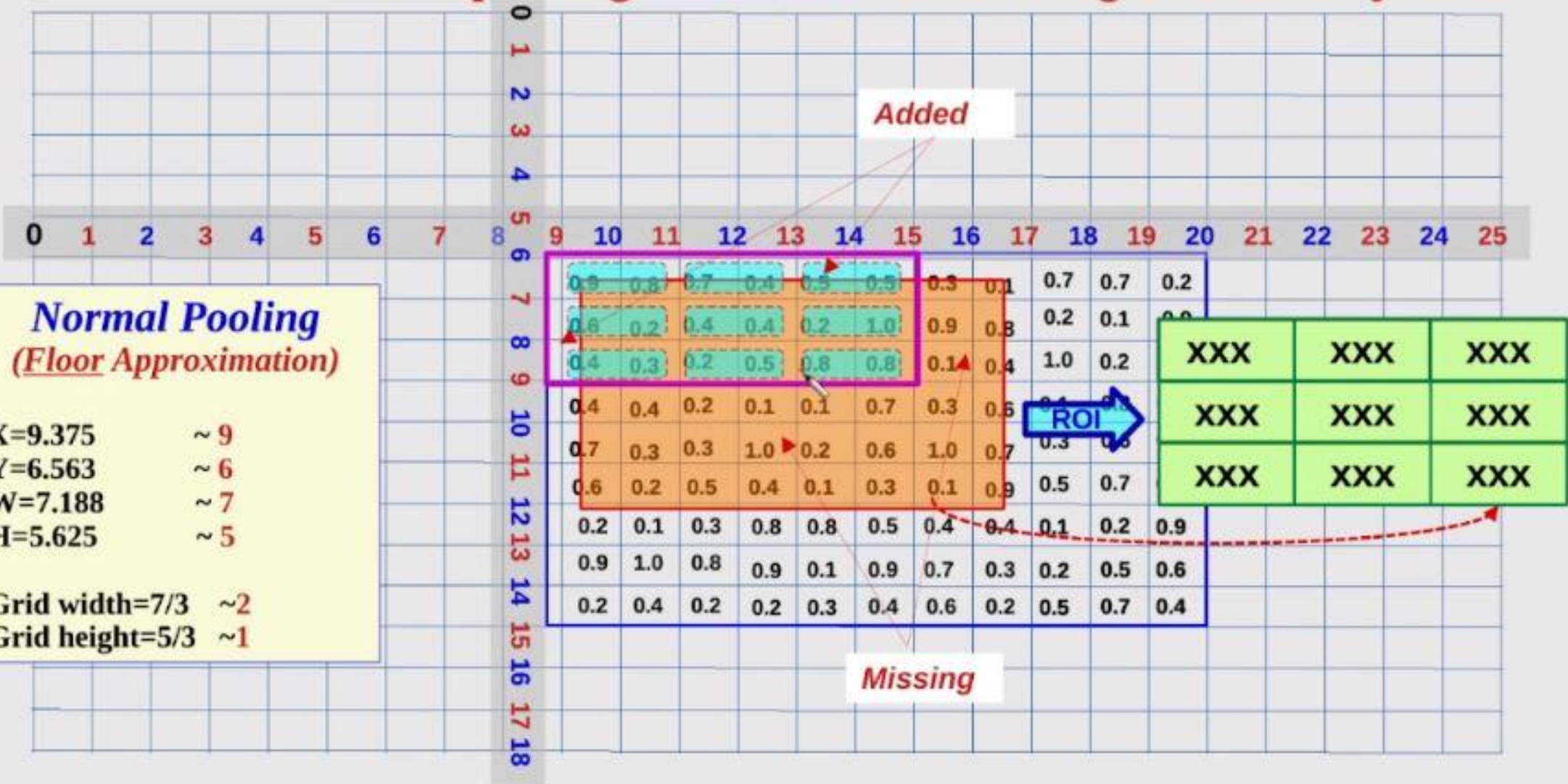
Remember:

Corresponding Traditional ROI Pooling for same Object

Normal Pooling
(Floor Approximation)

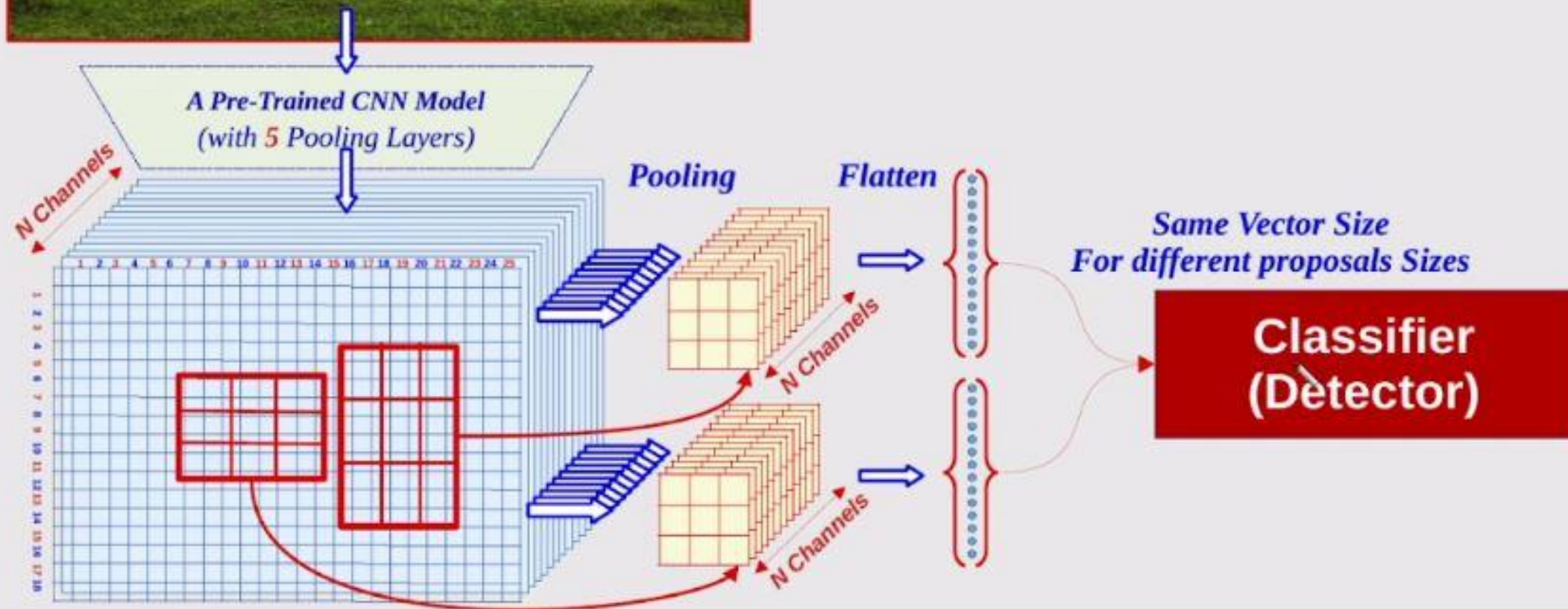
X=9.375 ~ 9
Y=6.563 ~ 6
W=7.188 ~ 7
H=5.625 ~ 5

Grid width=7/3 ~2
Grid height=5/3 ~1





*Max – Pooling is applied on Channel Basis
(repeat pooling for Each Channel)*



Architecture of Detector Network (for Fast RCNN, Faster RCNN)

123

