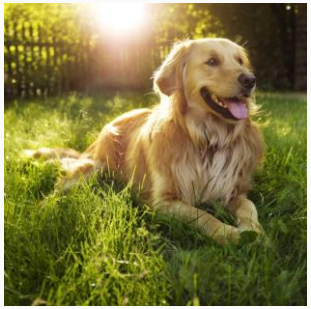
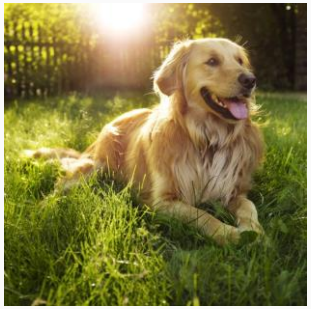
Lecture 7 : object detection and YOLO v3

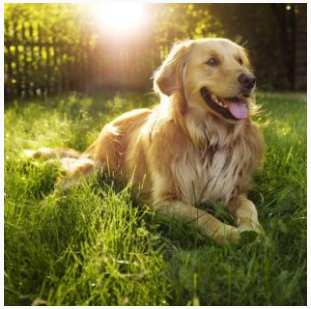
By : Khalil idrissi





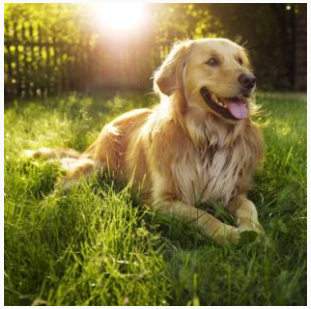
dog

classification

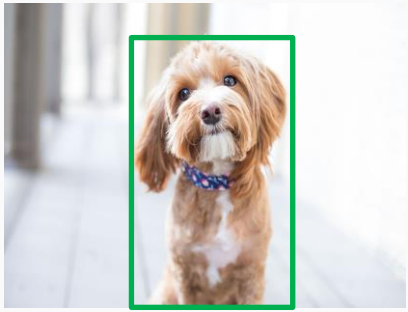


dog

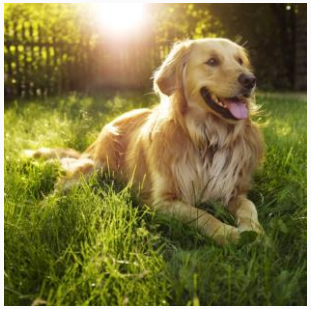
classification



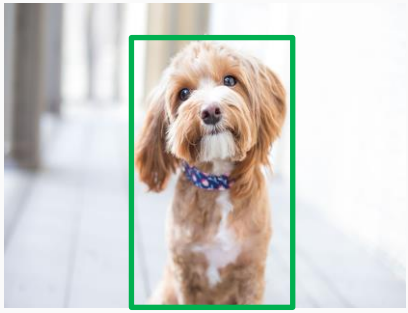
dog



classification

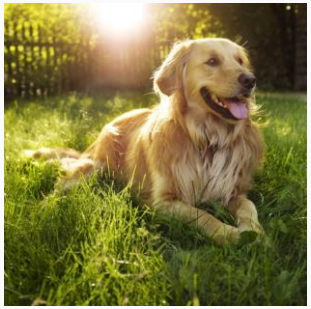


dog



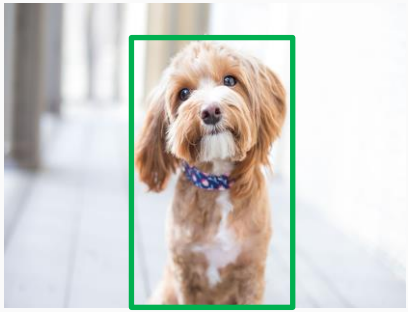
dog

classification



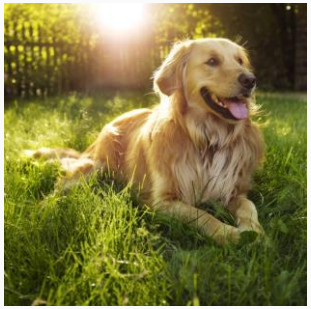
dog

Classification
+ localization



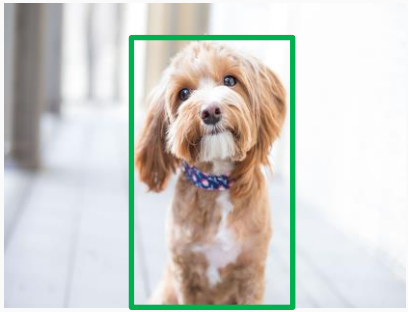
dog

classification

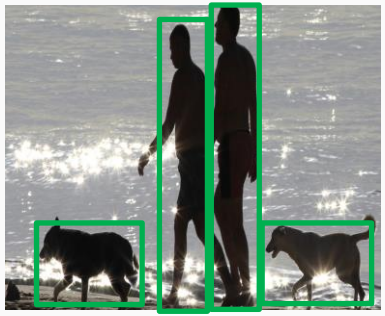


dog

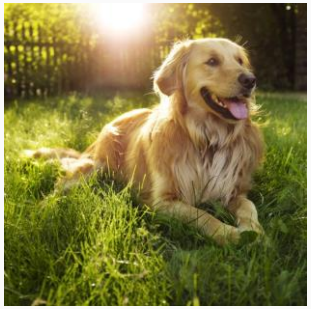
Classification
+ localization



dog

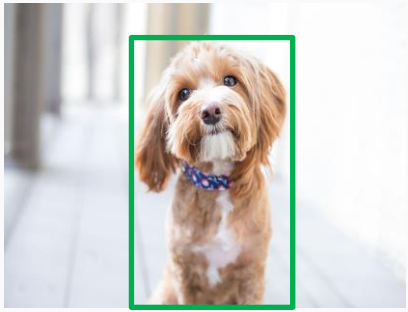


classification



dog

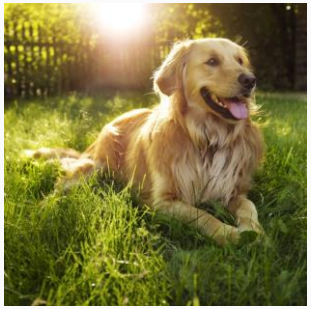
Classification
+ localization



dog

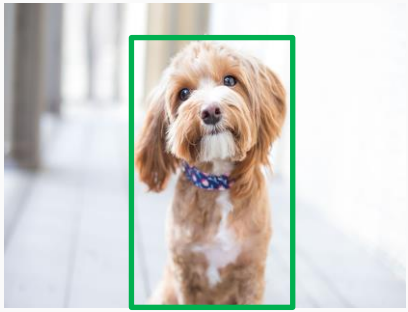
Dog, human

classification



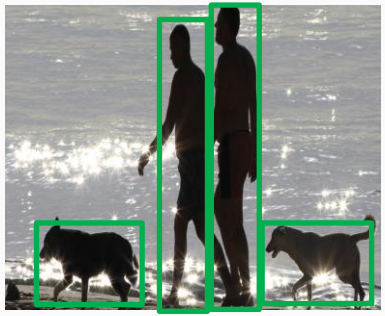
dog

Classification
+ localization



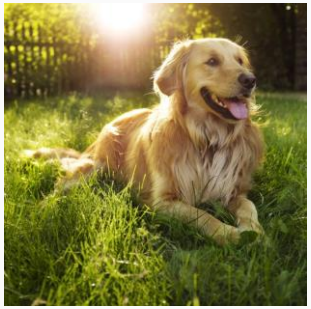
dog

Object detection



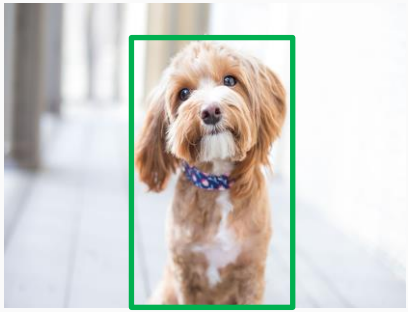
Dog, human

classification



dog

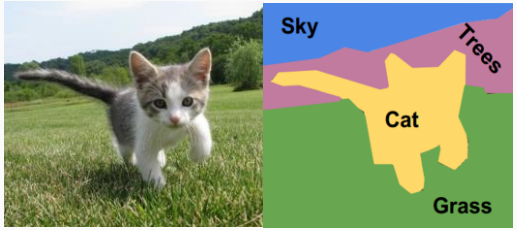
Classification
+ localization



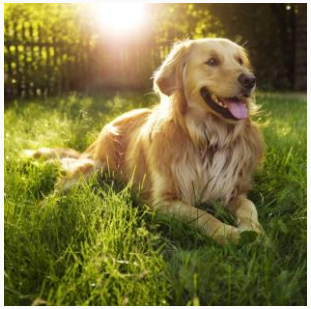
dog

Object detection

Dog, human

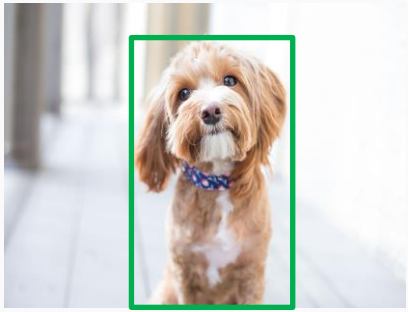


classification



dog

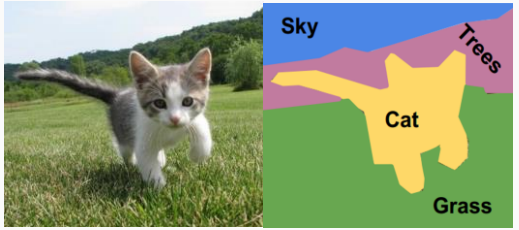
Classification
+ localization



dog

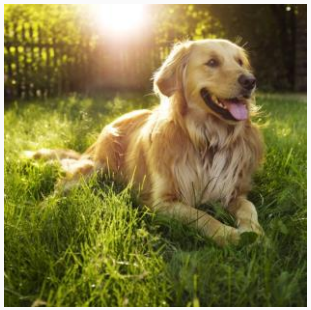
Object detection

Dog, human



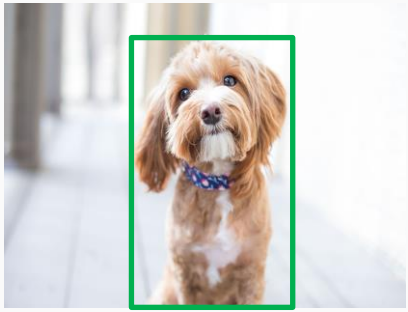
Cat, sky, grass, trees

classification



dog

Classification
+ localization

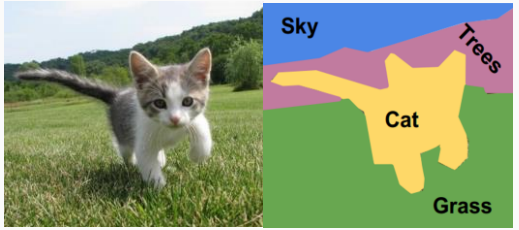


dog

Object detection

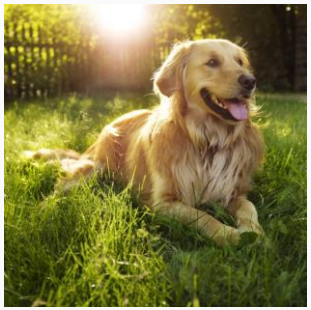
Dog, human

Semantic segmentation



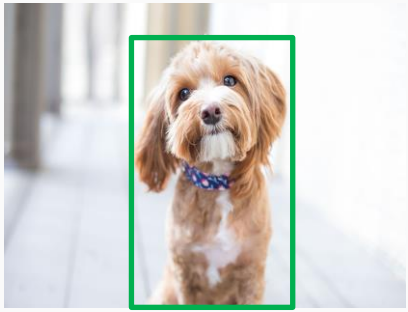
Cat, sky, grass, trees

classification



dog

Classification
+ localization

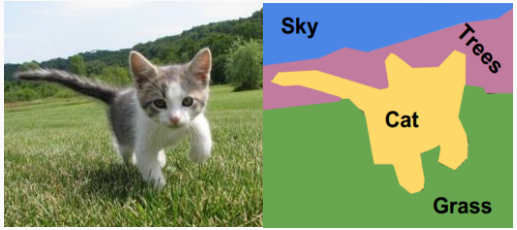


dog

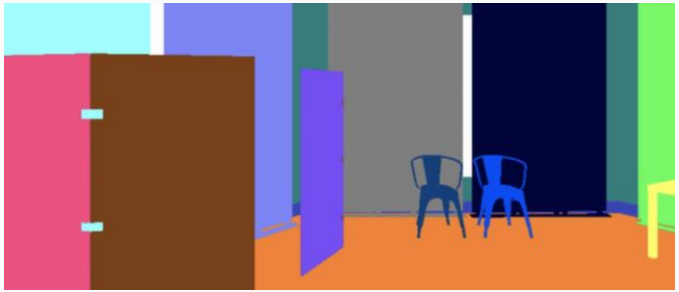
Object detection

Dog, human

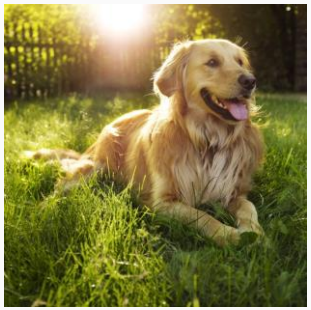
Semantic segmentation



Cat, sky, grass, trees

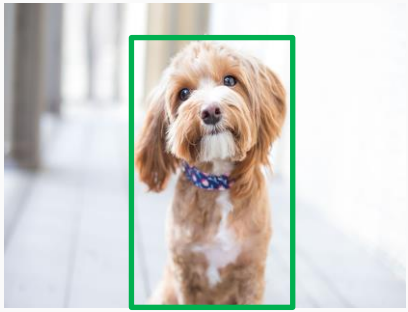


classification



dog

Classification
+ localization

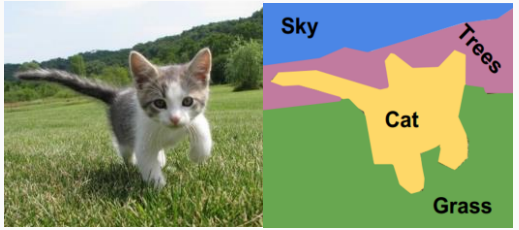


dog

Object detection

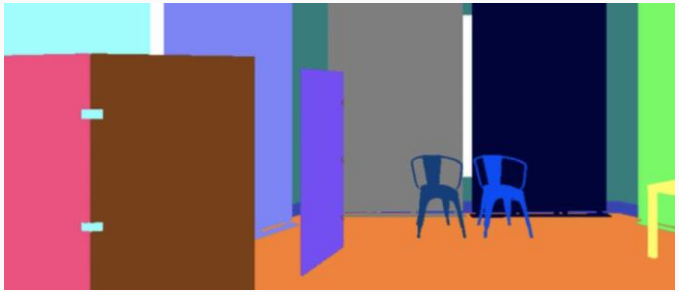
Dog, human

Semantic segmentation

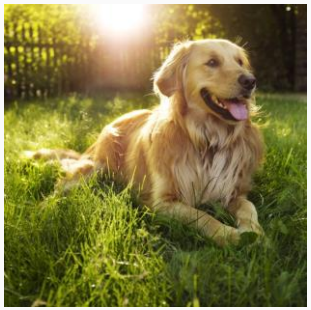


Cat, sky, grass, trees

instance segmentation

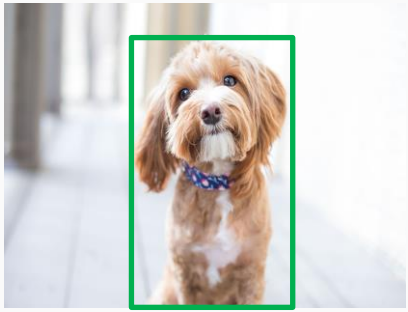


classification



dog

Classification
+ localization

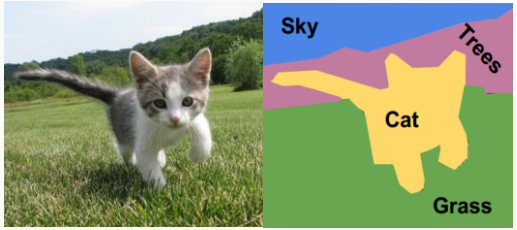


dog

Object detection

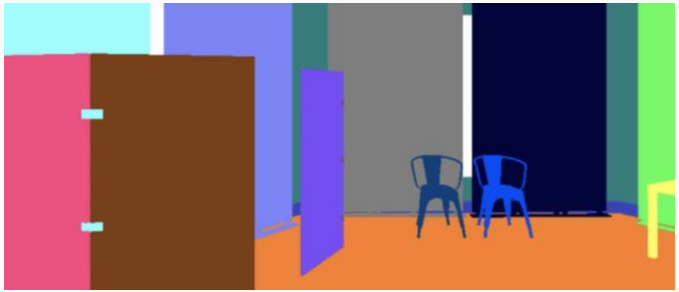
Dog, human

Semantic segmentation

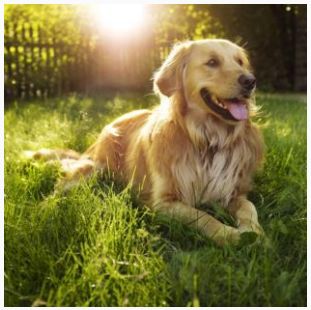


Cat, sky, grass, trees

instance segmentation

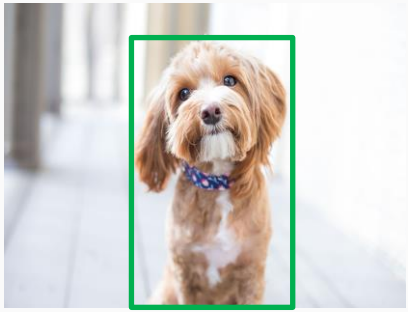


classification



dog

Classification
+ localization



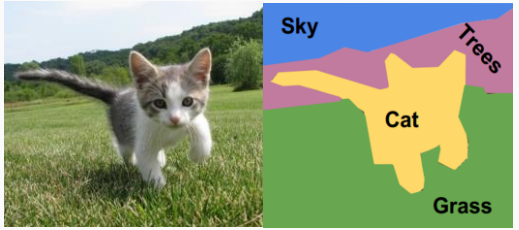
dog

Single object

Object detection

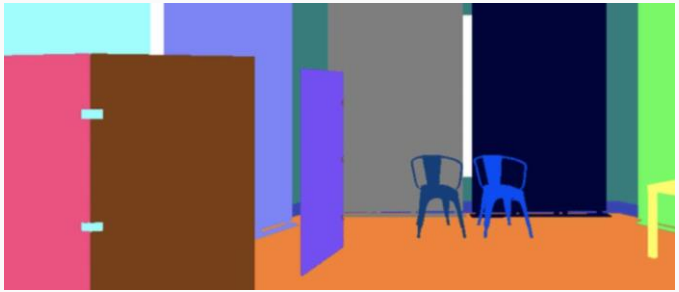
Dog, human

Semantic segmentation

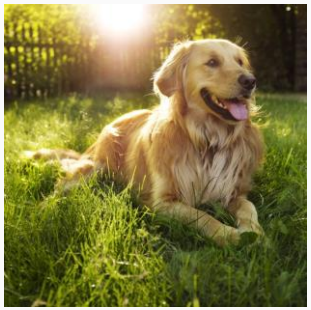


Cat, sky, grass, trees

instance segmentation

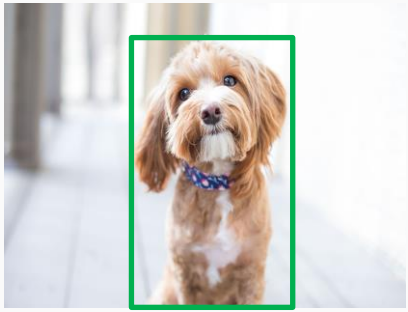


classification



dog

Classification
+ localization



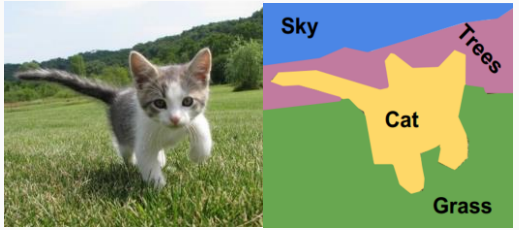
dog

Single object

Object detection

Dog, human

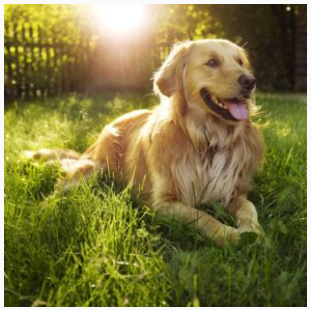
Semantic segmentation



Cat, sky, grass, trees

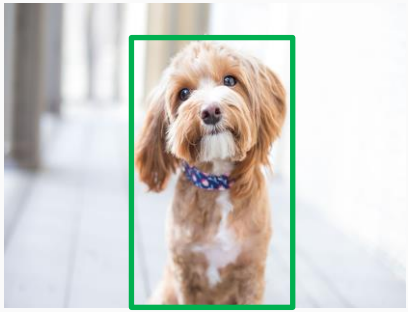
instance segmentation

classification



dog

Classification
+ localization



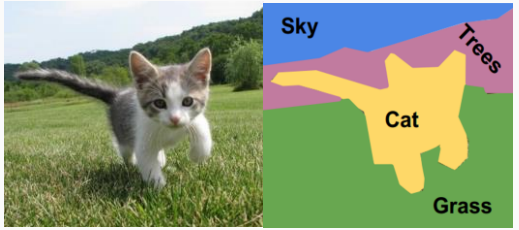
dog

Single object

Object detection

Dog, human

Semantic segmentation



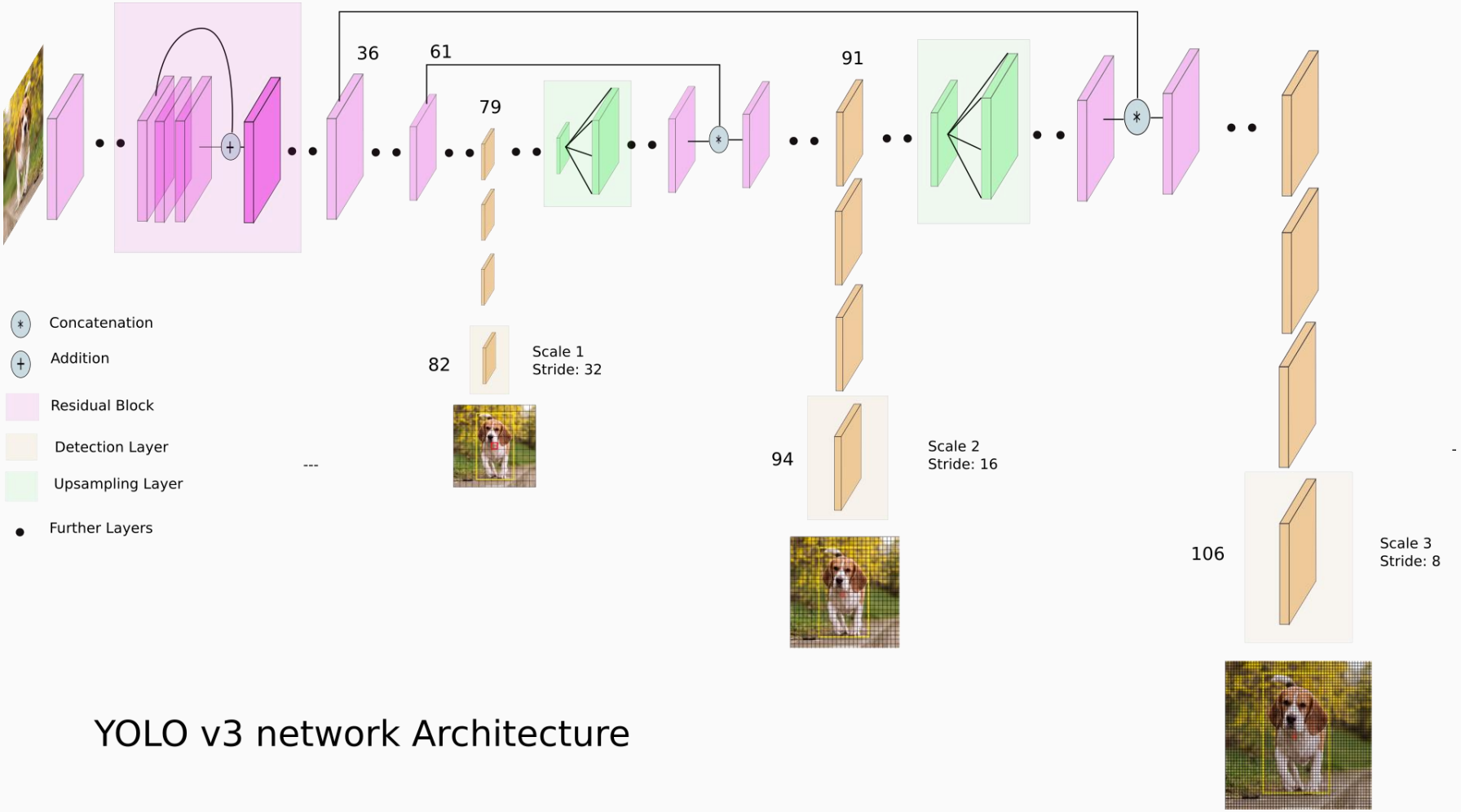
Cat, sky, grass, trees

multiple objects

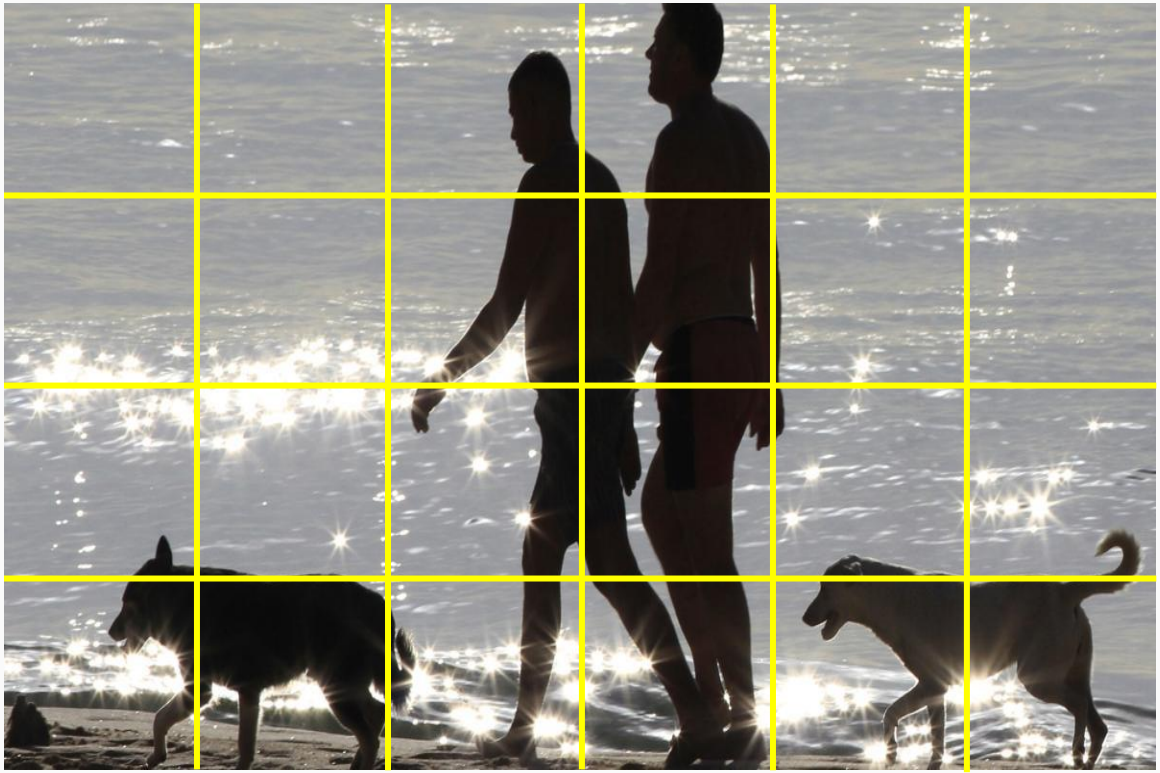
instance segmentation

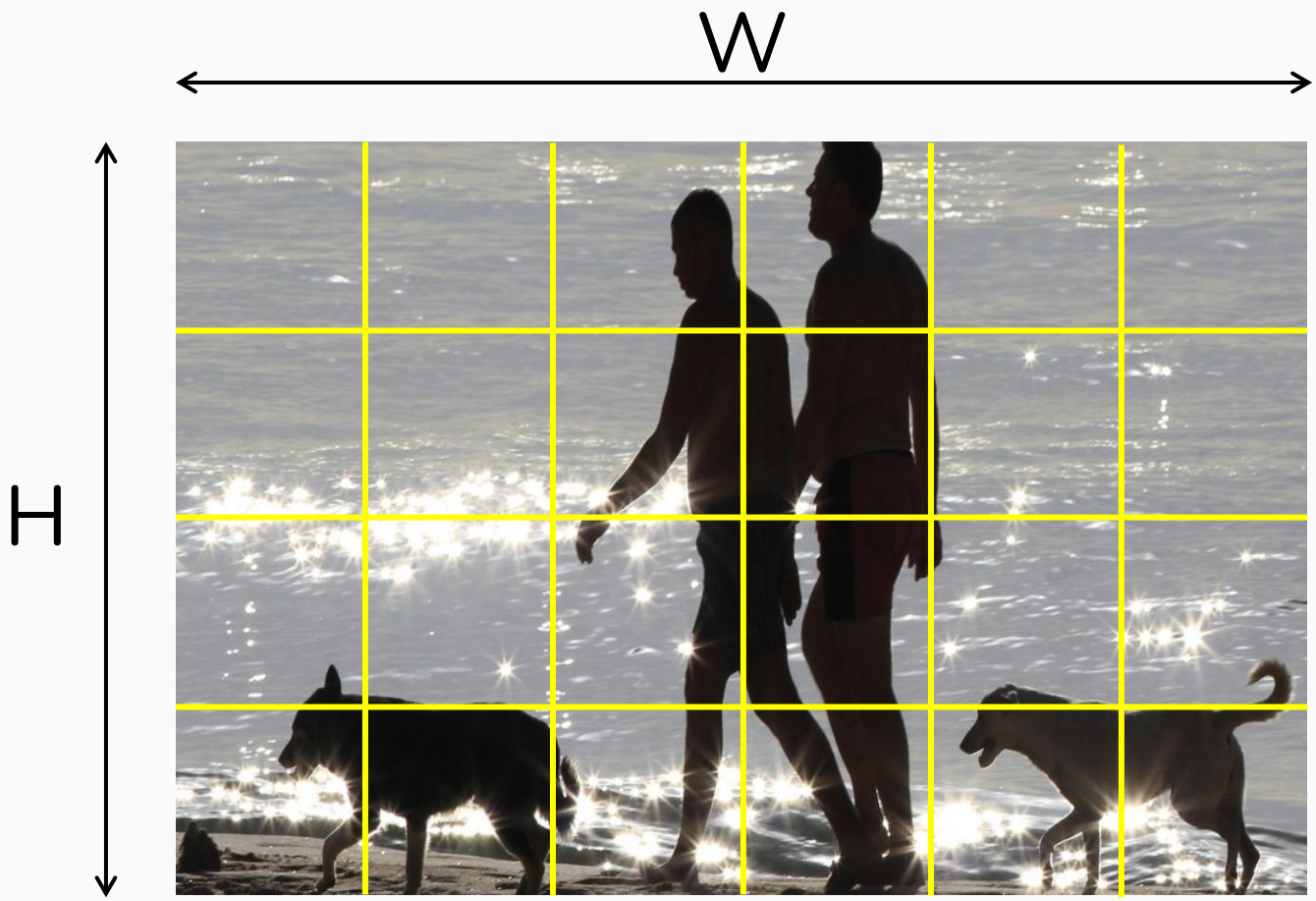
Object detection

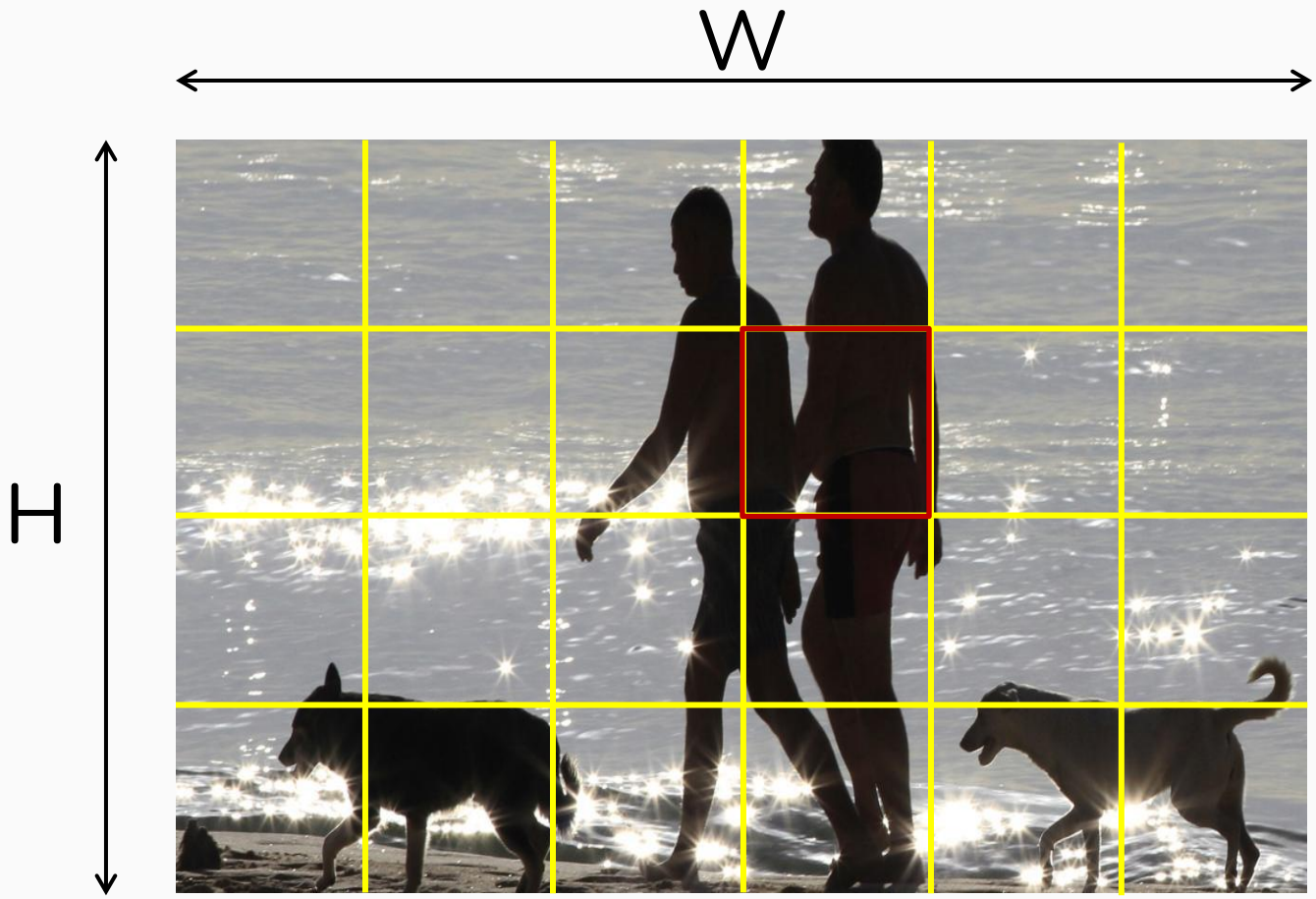
Dog, human

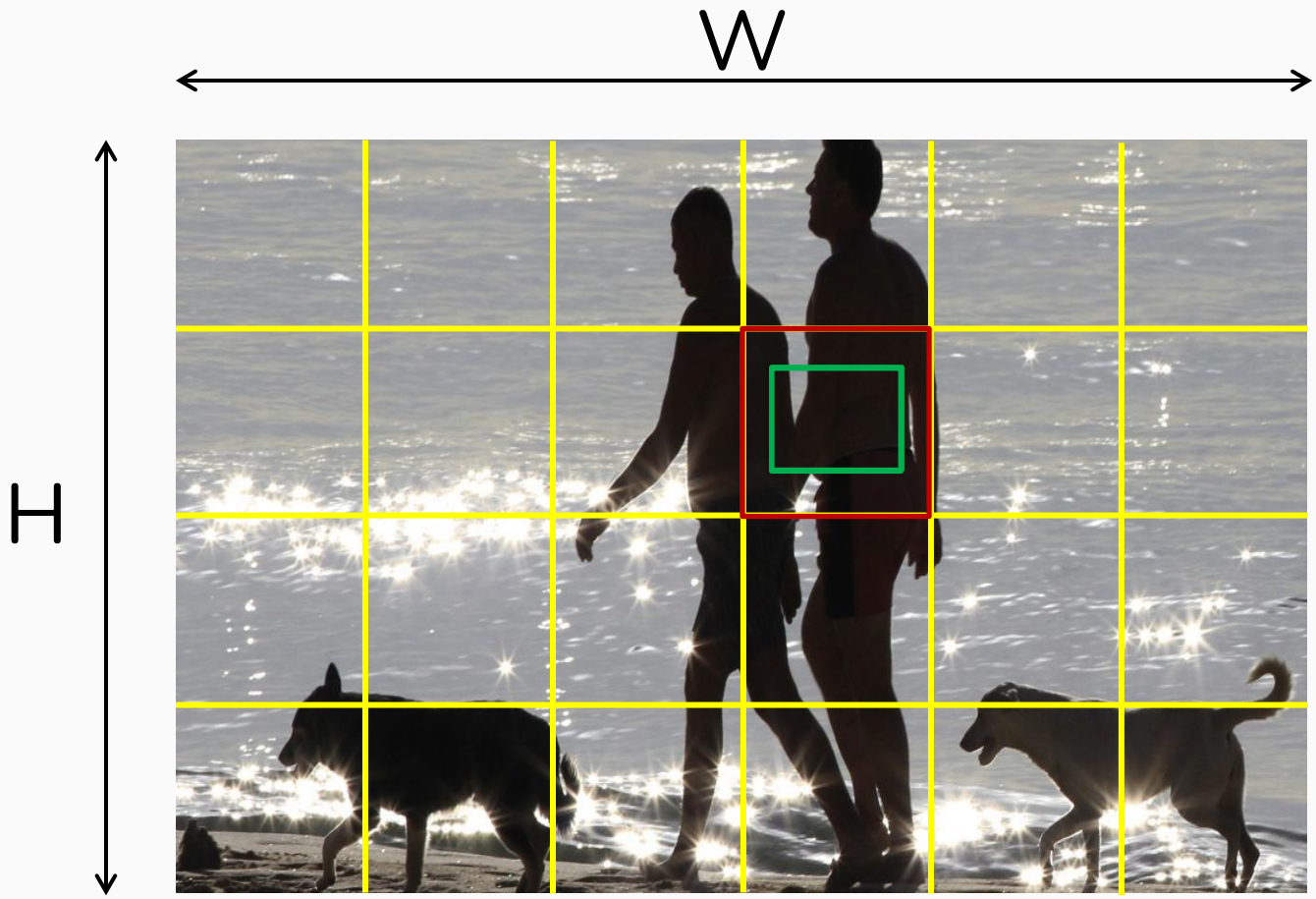


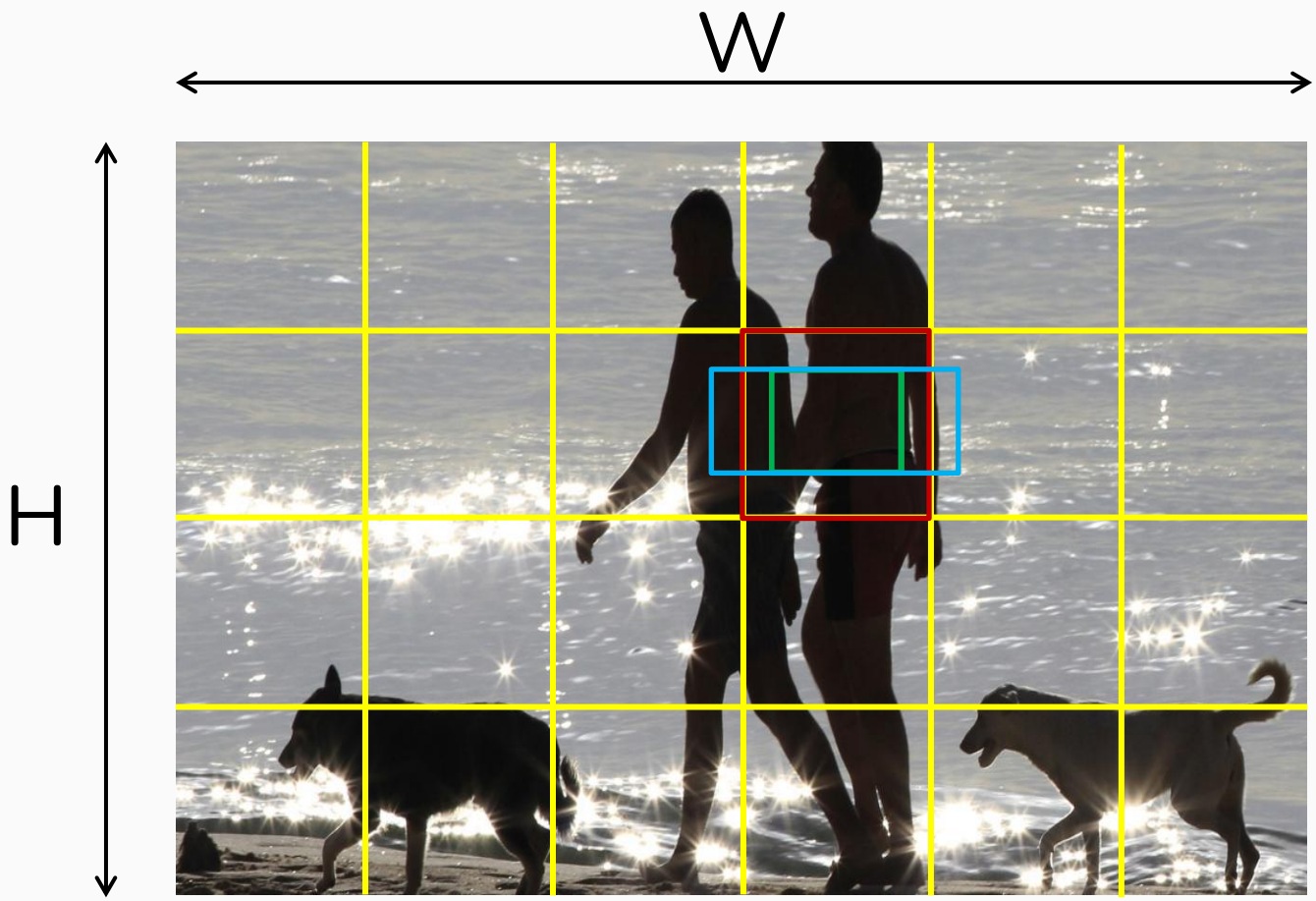
YOLO v3 network Architecture

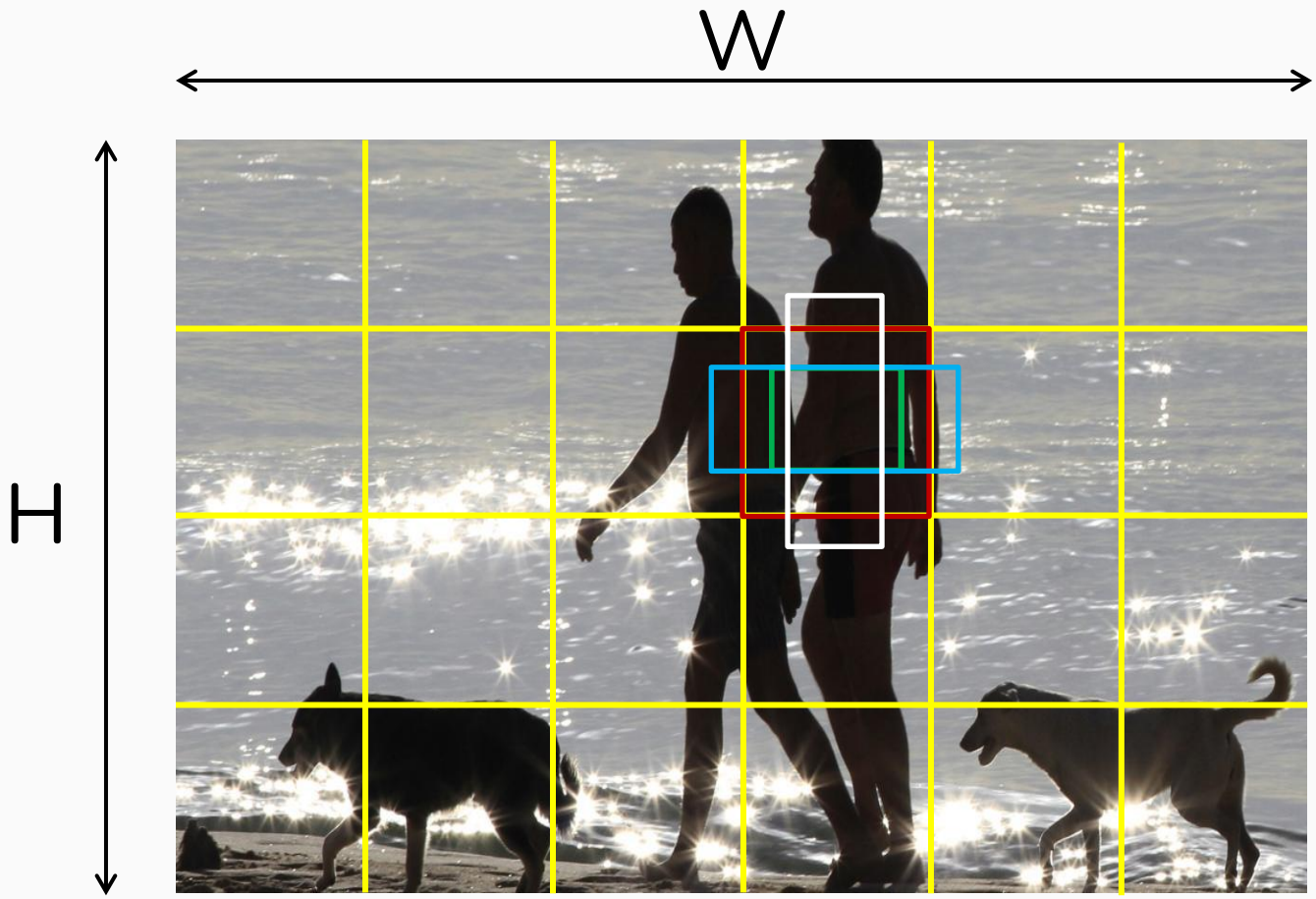


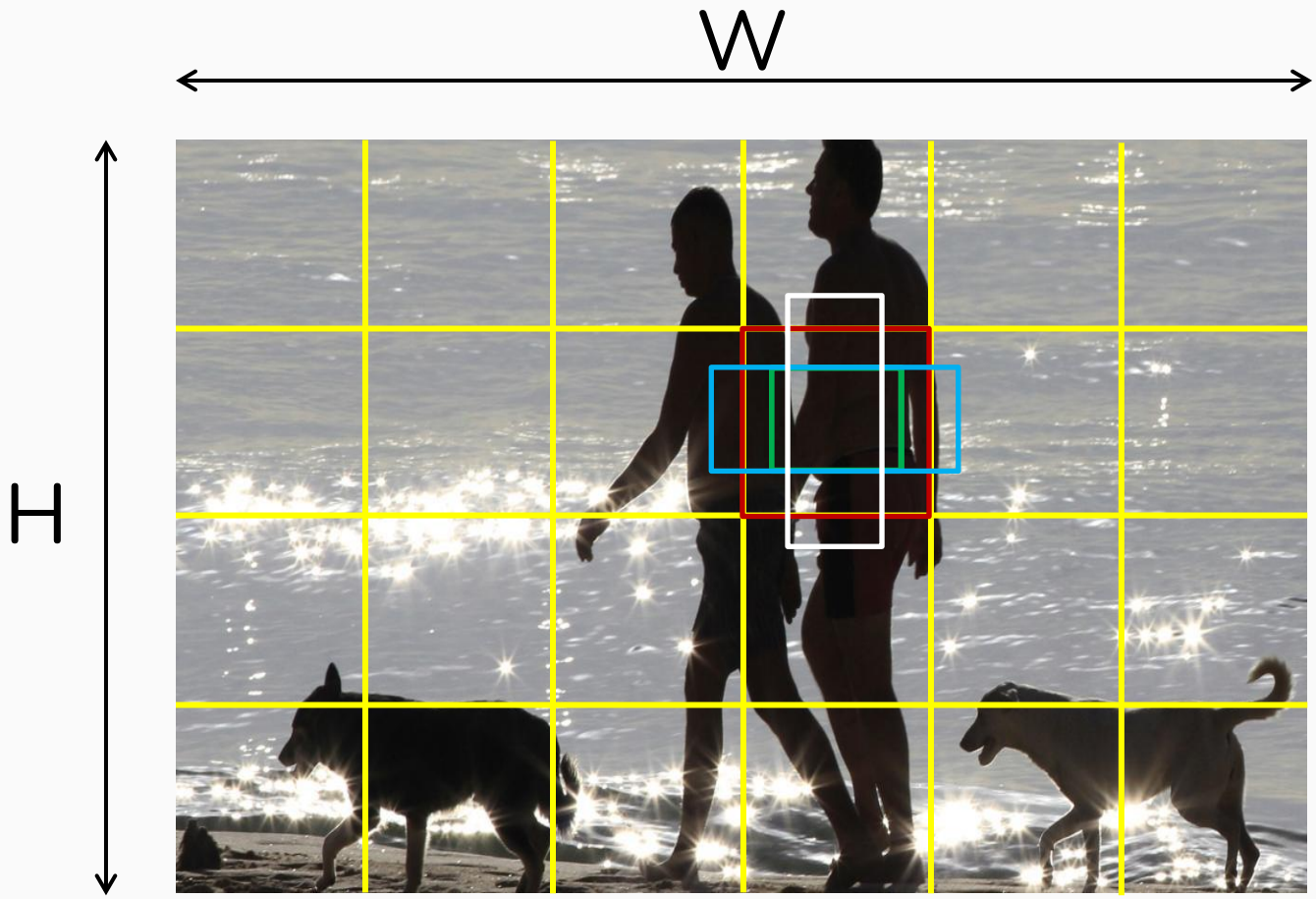

W
H

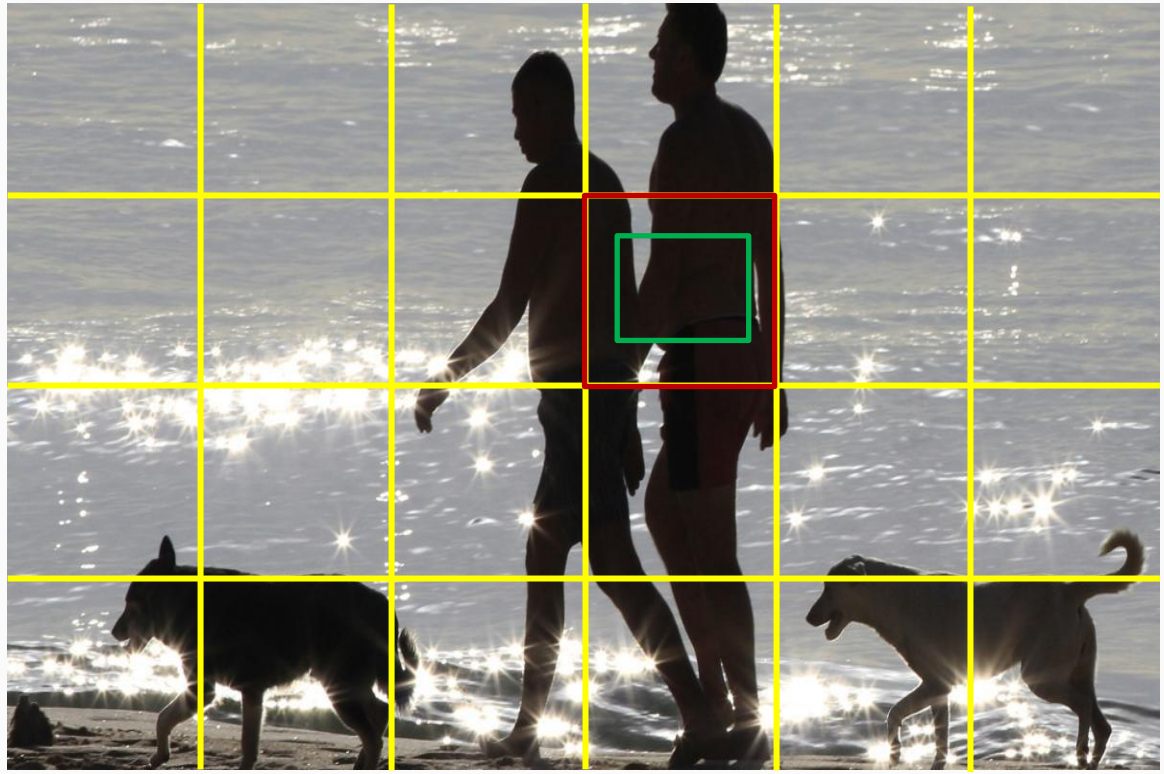

W
H


W
H

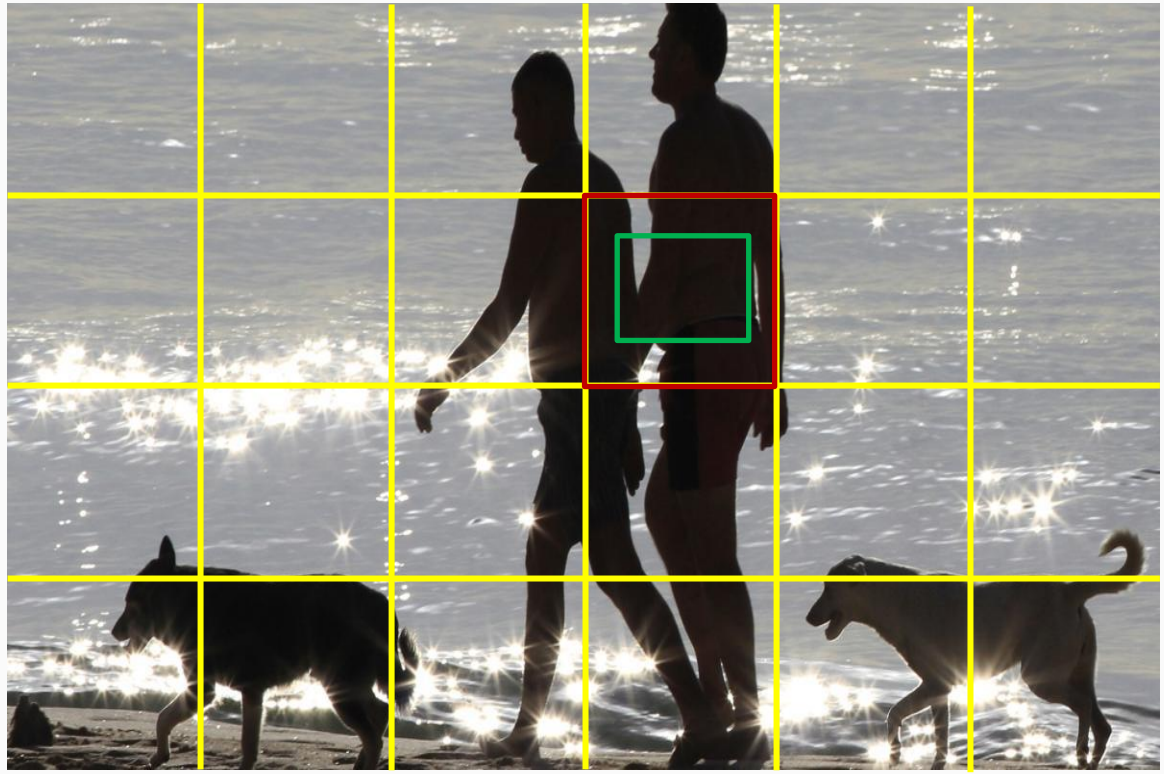

W
H


W
H

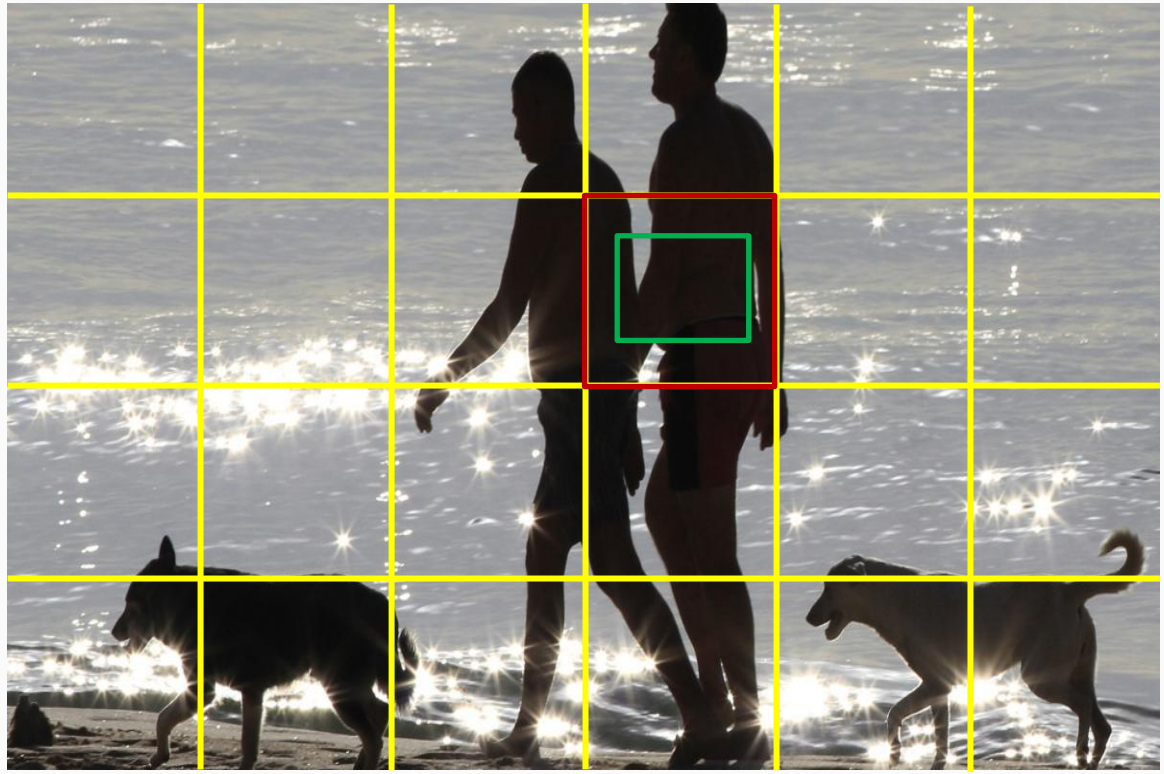

W
H



$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$

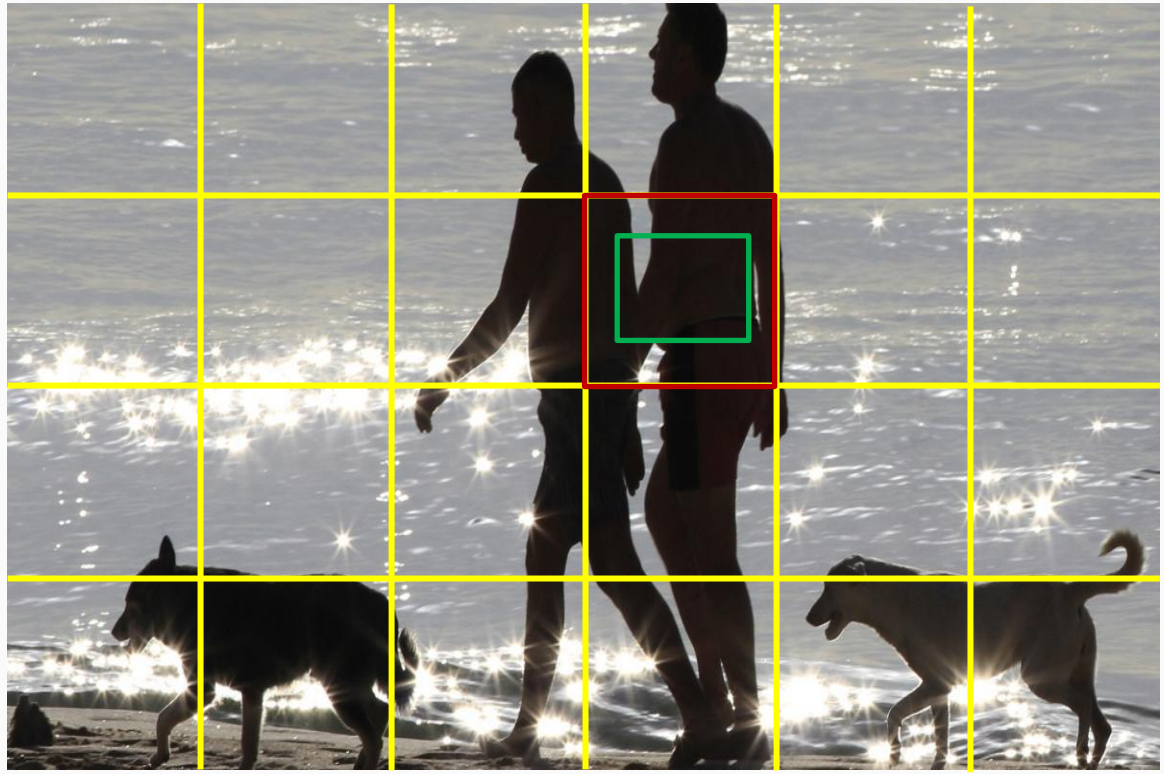


$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$

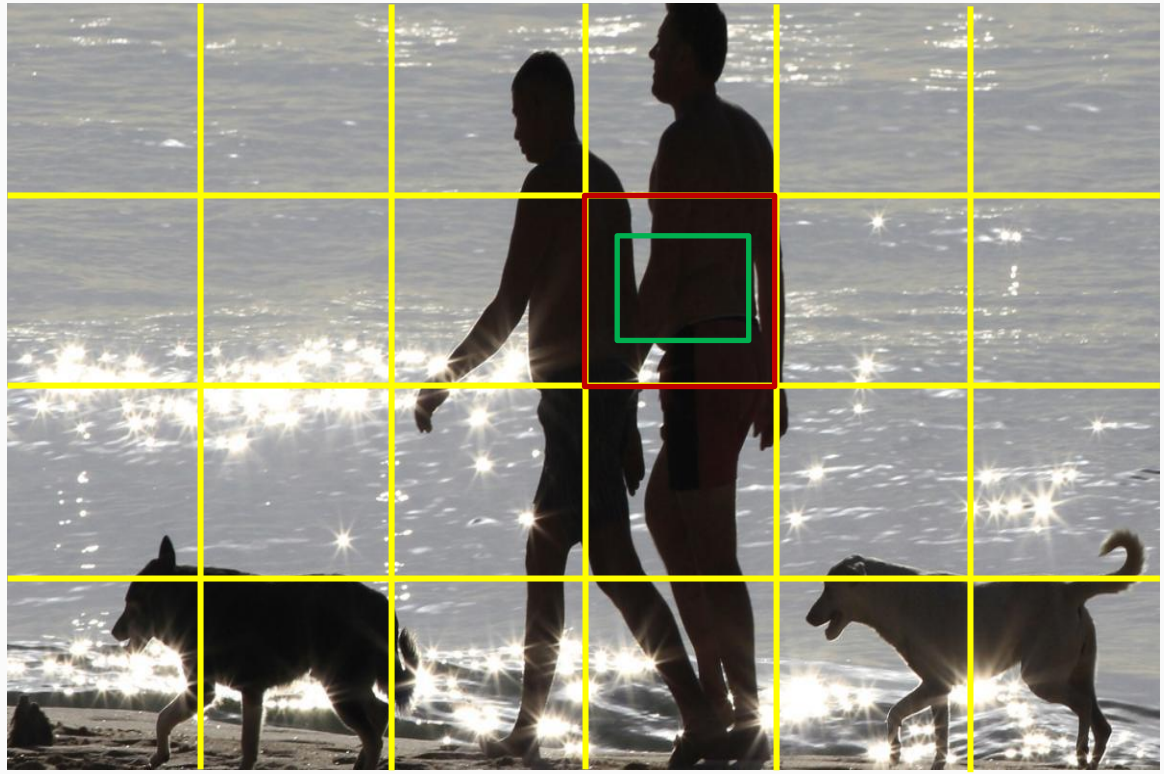


Confidence score that an object is in the box

$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$

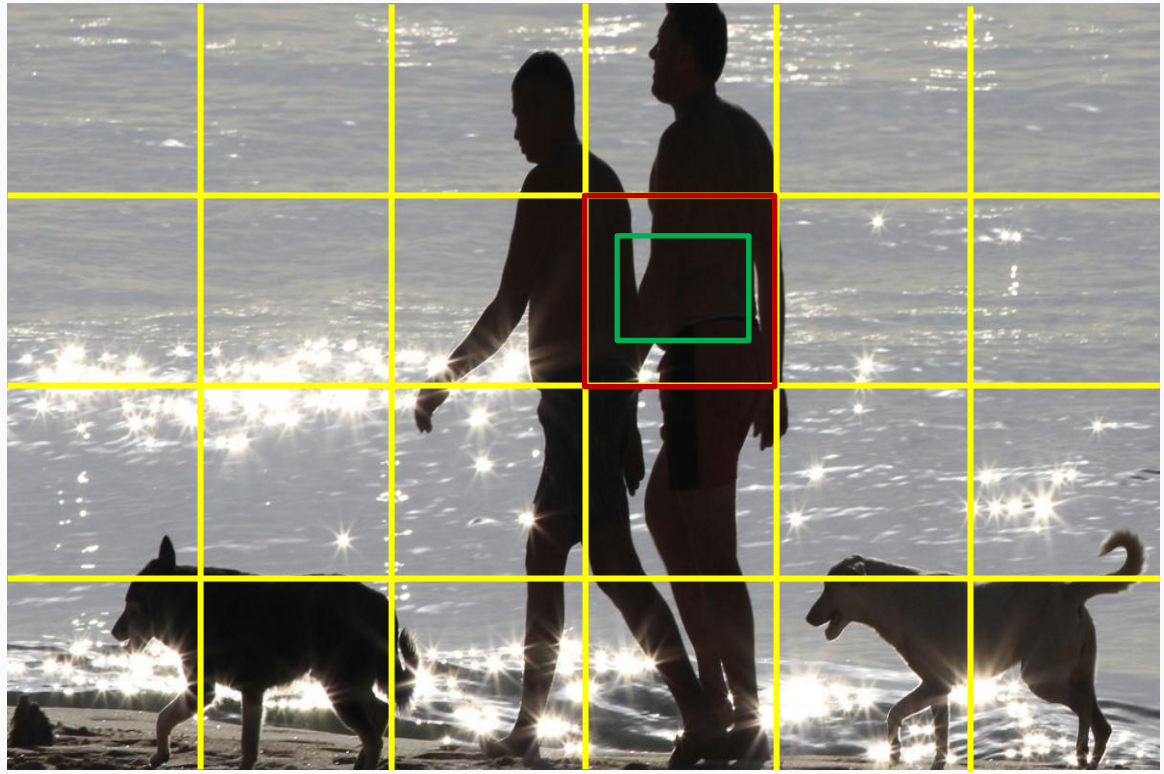


$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$

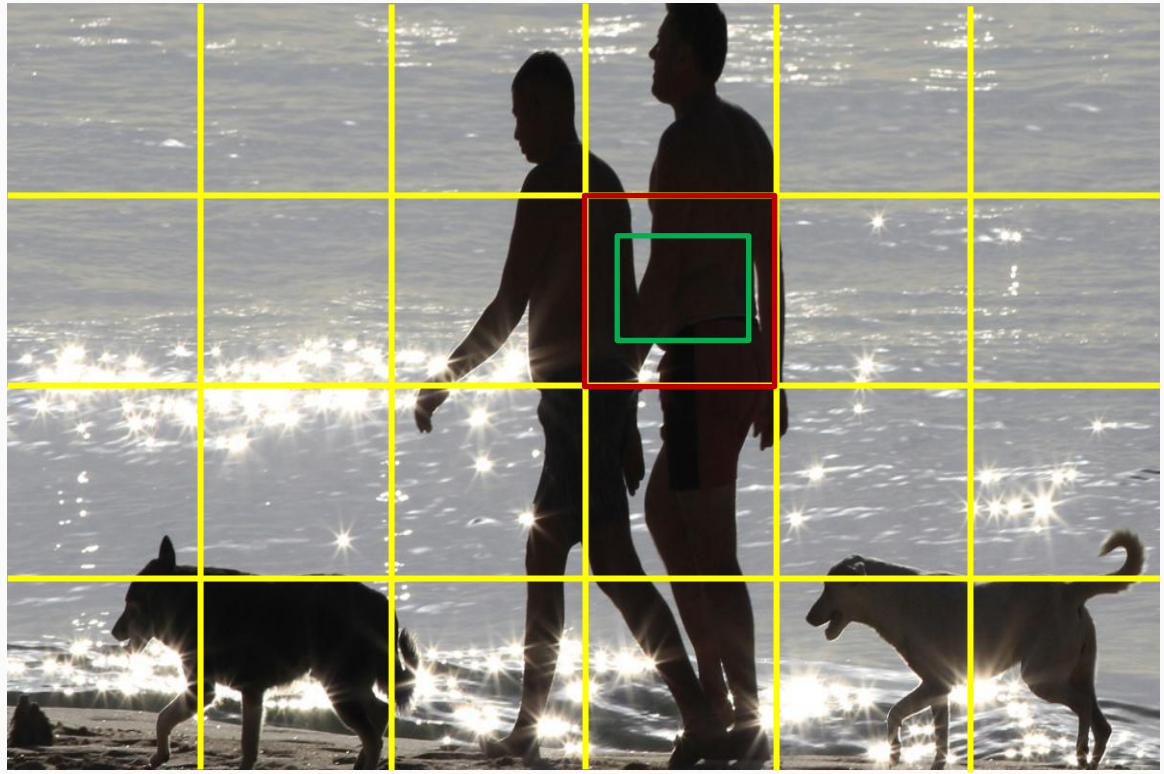


The coordinates of the center of the box, width, height

$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$

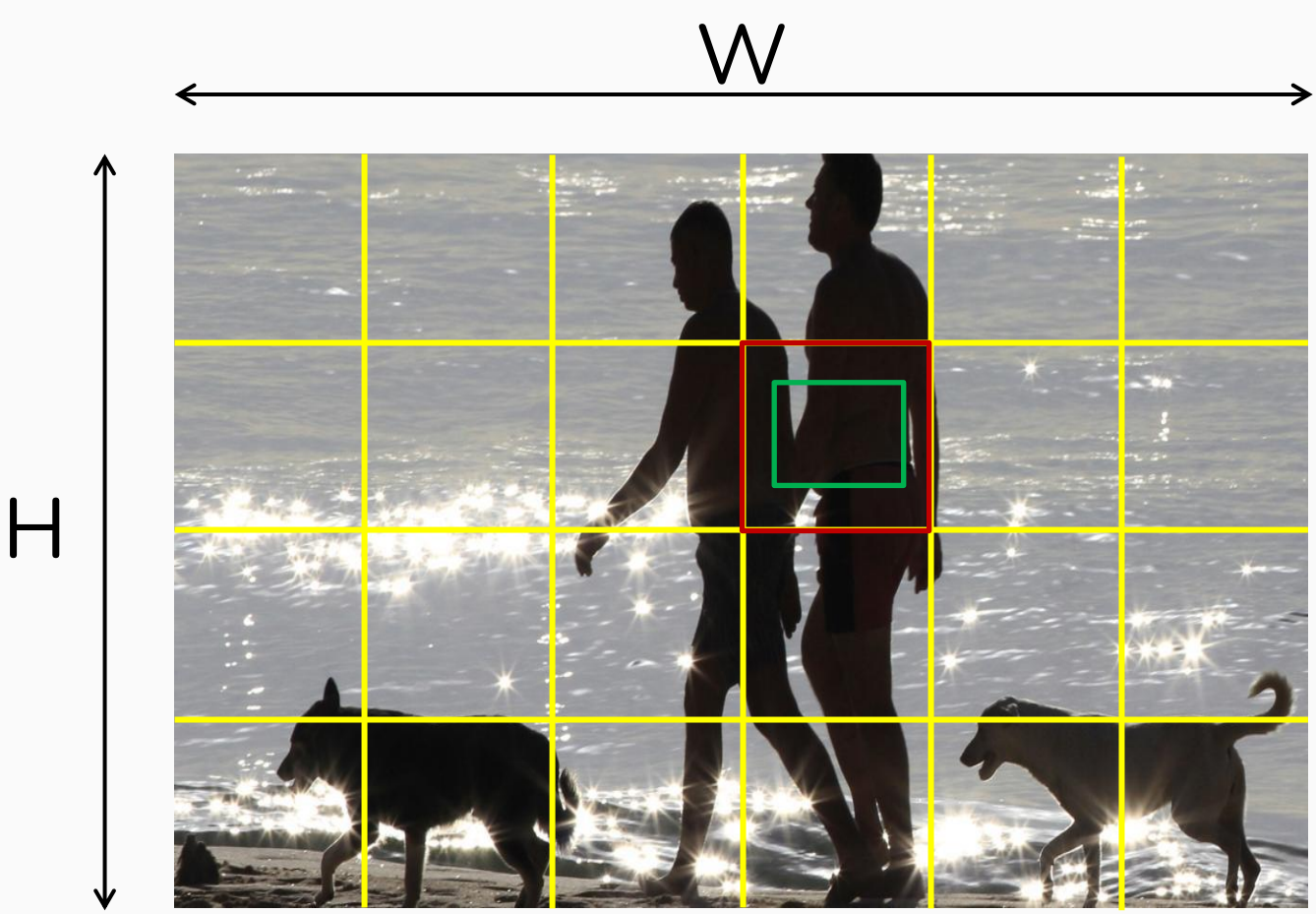


$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$

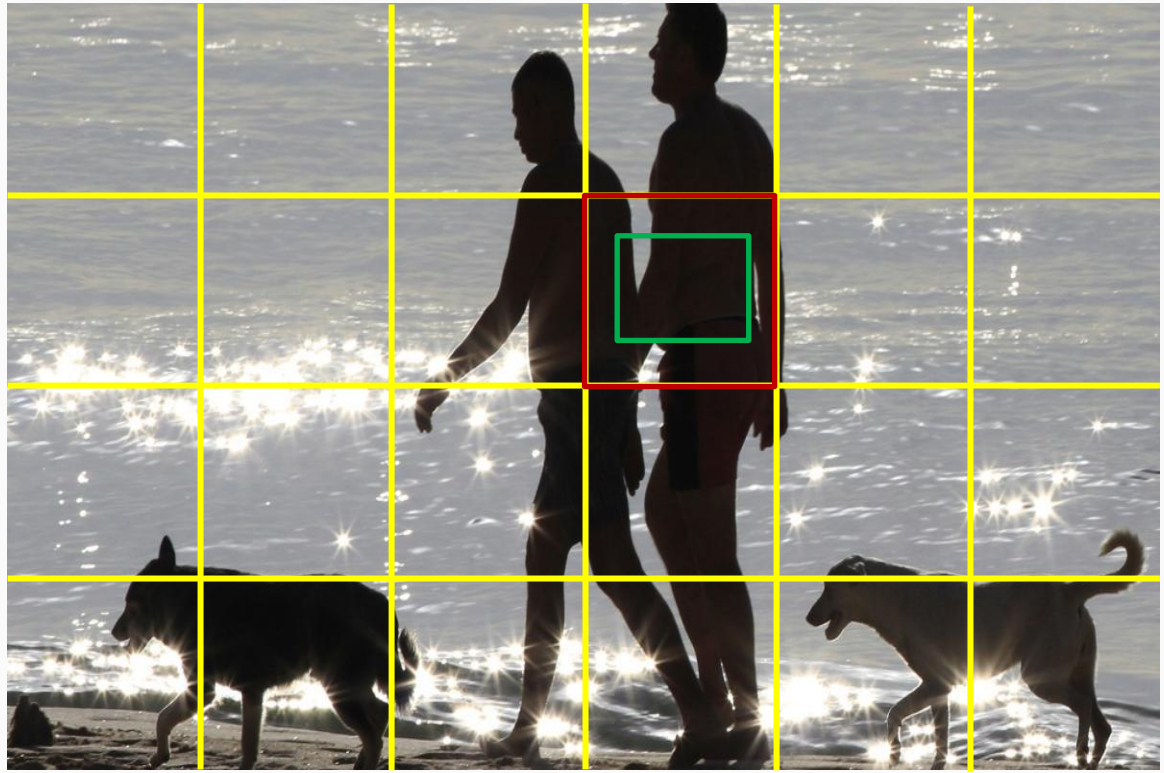


Probability that this box contains an object 1,2

$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$



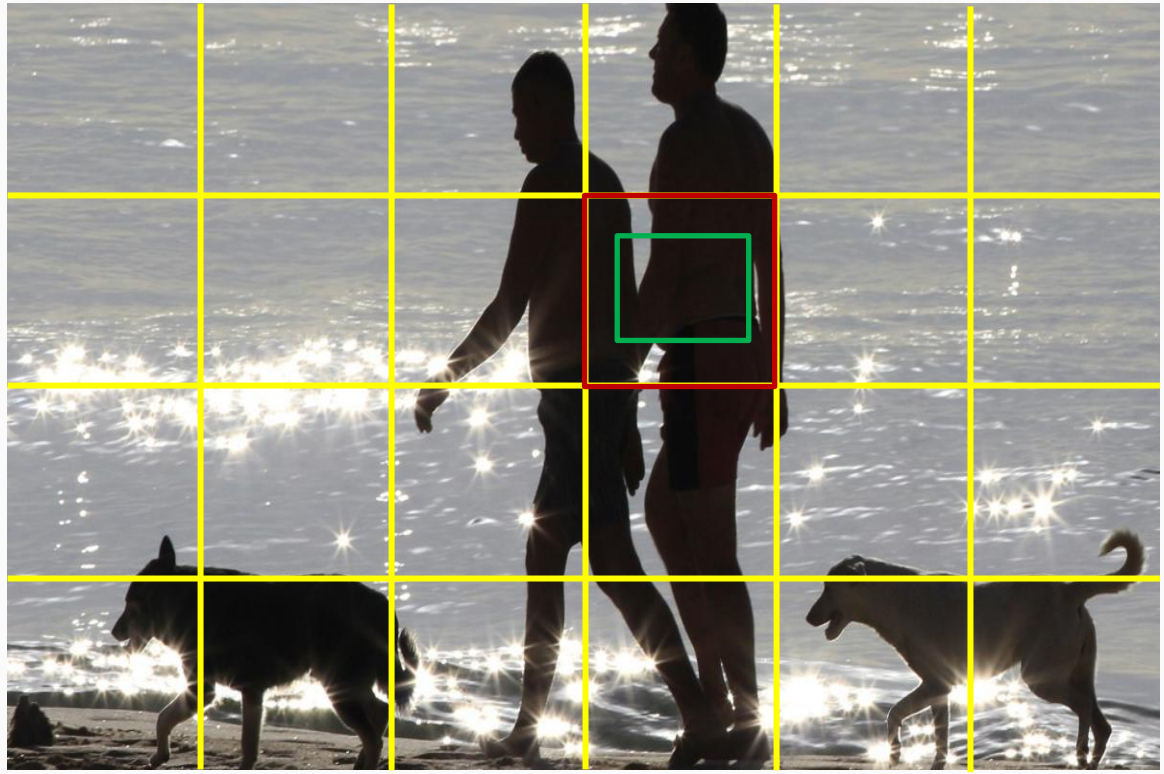
Refining the anchor boxes



Refining the anchor boxes

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

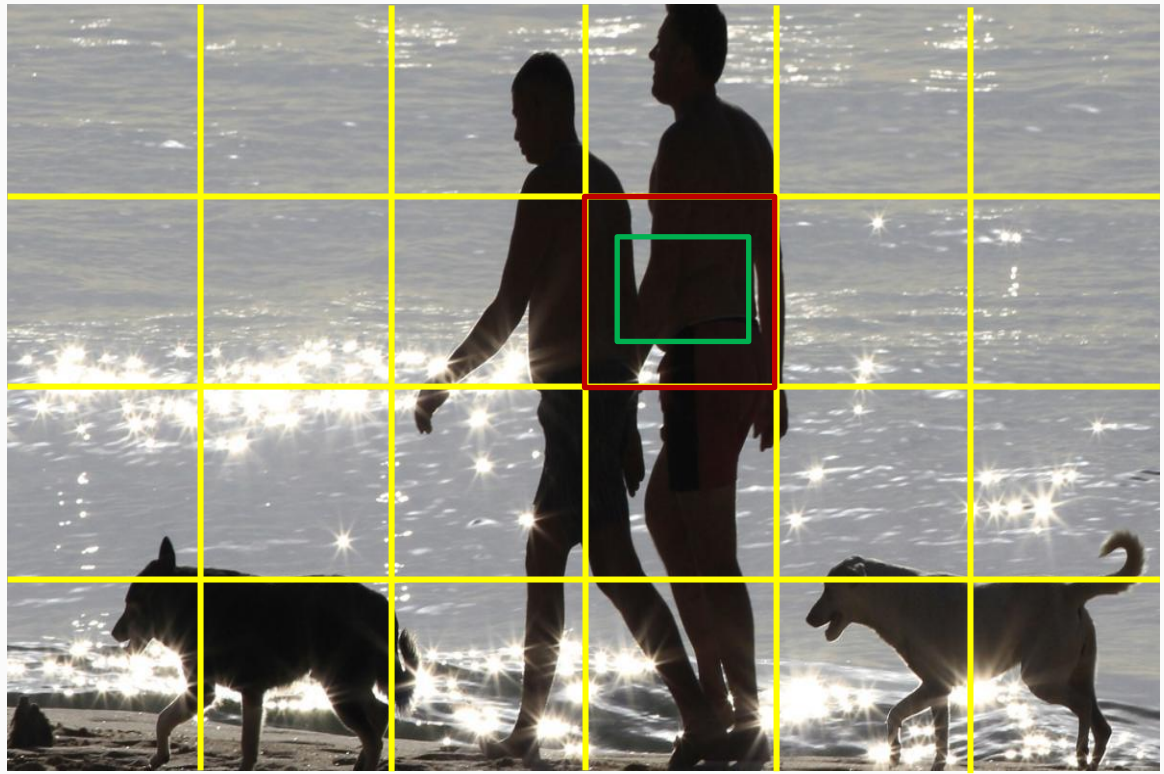
t_x, t_y, t_h, t_w are the outputs from the last layer



Refining the anchor boxes

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

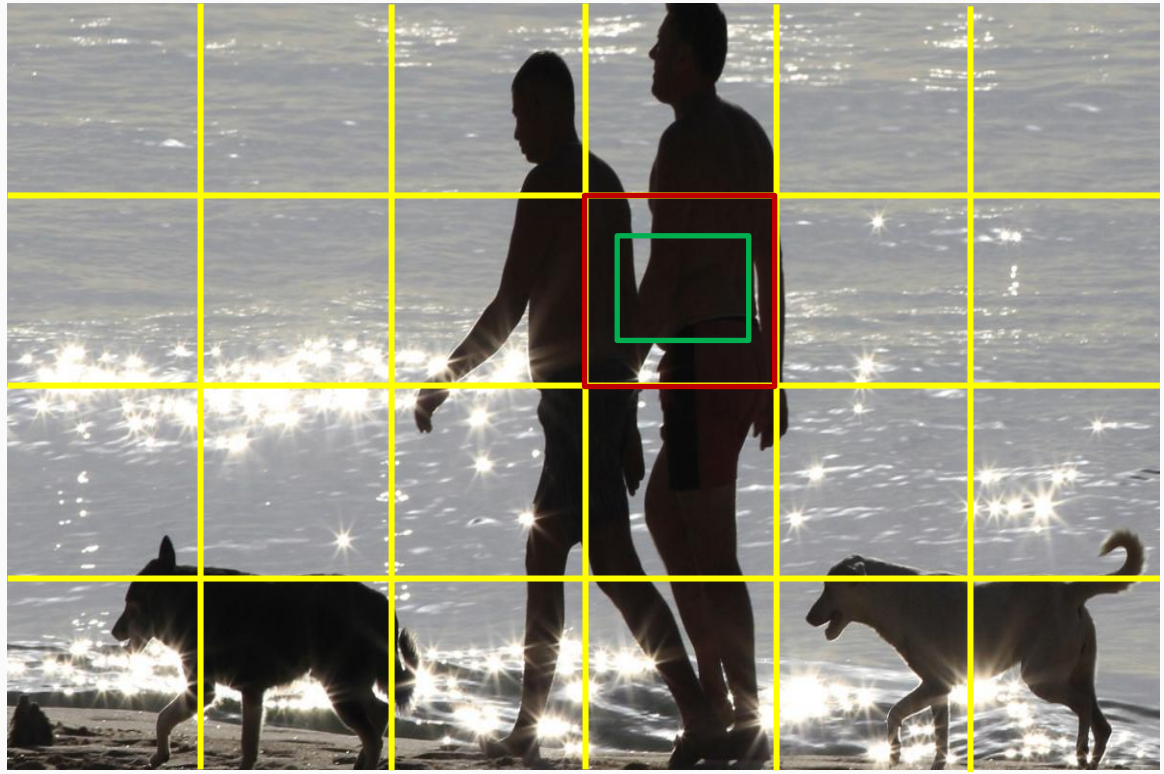
b_x, b_y, b_h, b_w are the coordinates for the refined bounding box



Refining the anchor boxes

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

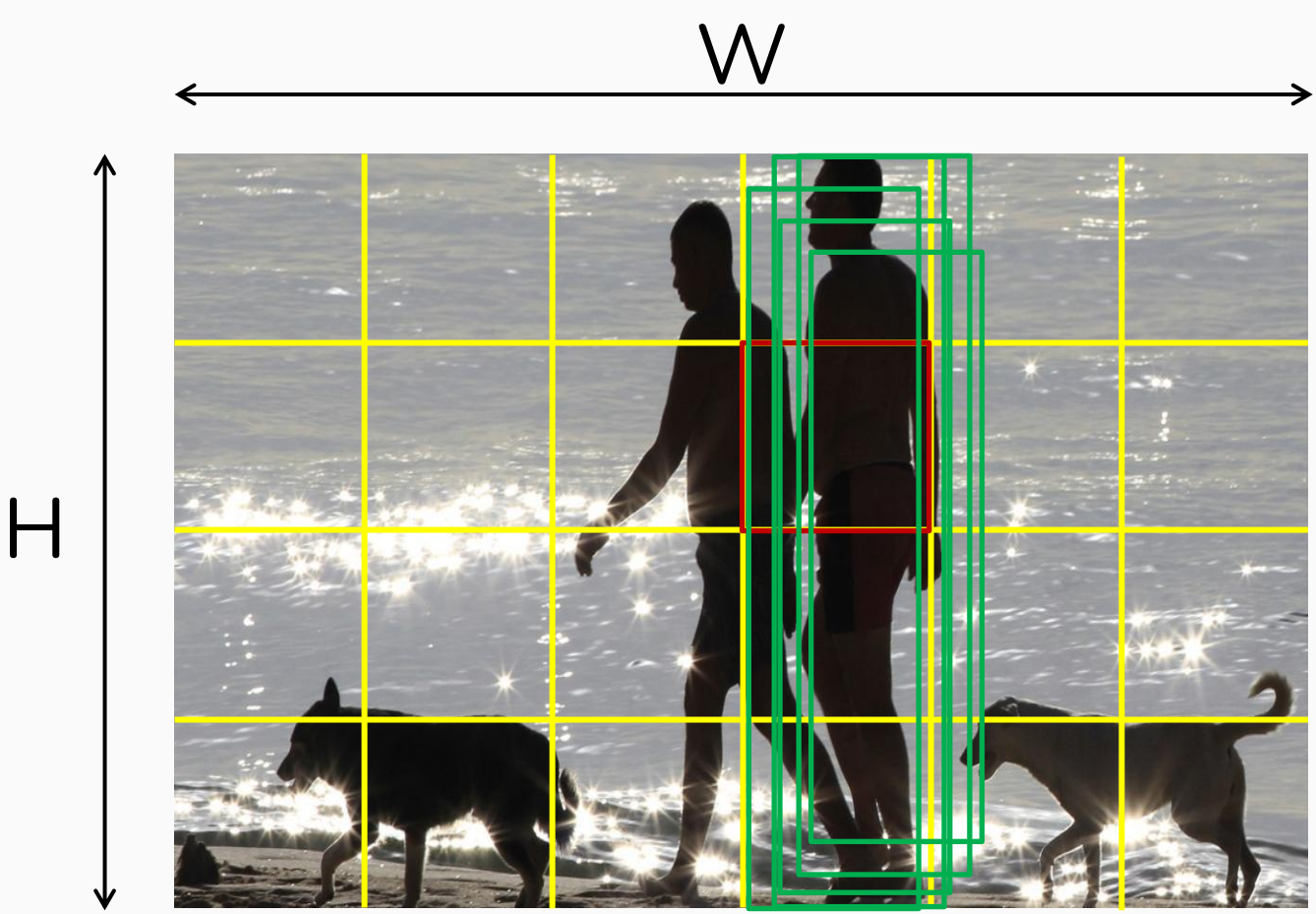
p_w , p_h are the original size if the anchor box



Refining the anchor boxes

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

c_x, c_y are coordinates of the current grid cell



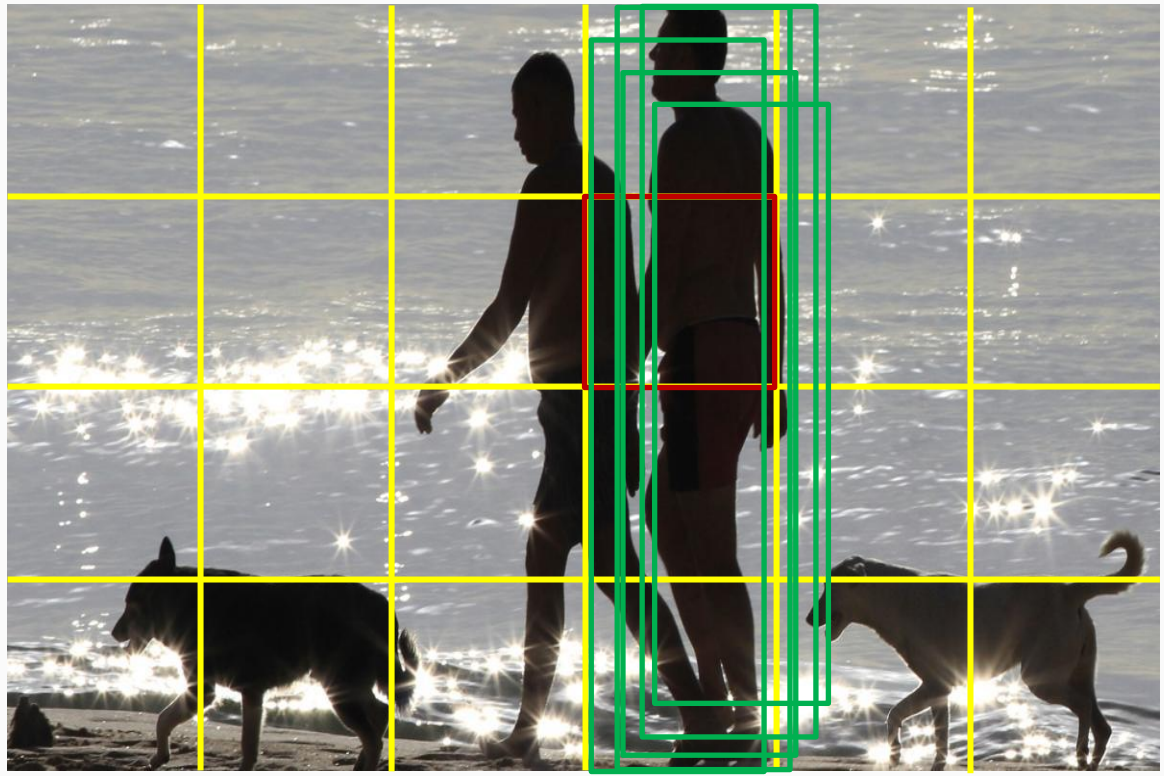
Refining the anchor boxes

$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$



Refining the anchor boxes

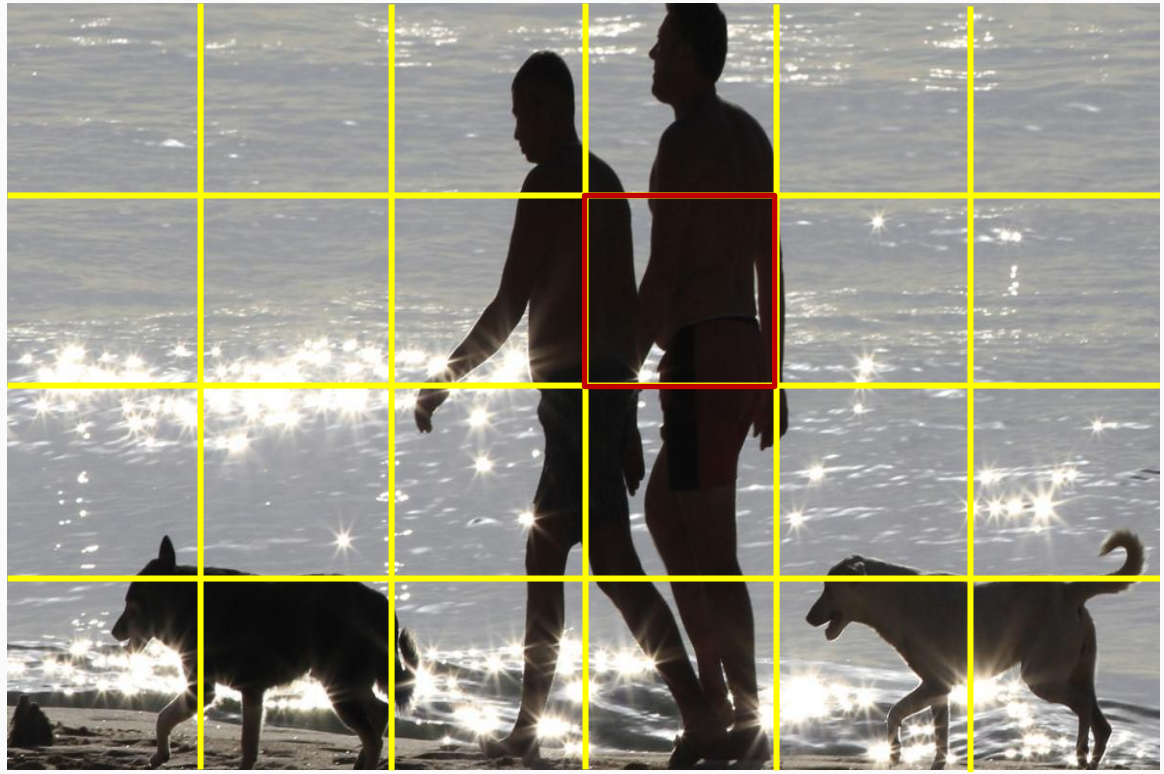
$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

NMS : non max suppression

Refining the anchor boxes

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

NMS : non max suppression



$$y = \begin{pmatrix} c_1 \\ t_x \\ t_y \\ t_h \\ t_w \\ p_1 \\ p_2 \end{pmatrix}$$

The core idea is to reframe the object detection as as single regression problem