**4. Introducción a *object detection* = *clasificación* + localización + seguimiento**

**Detección de objetos**

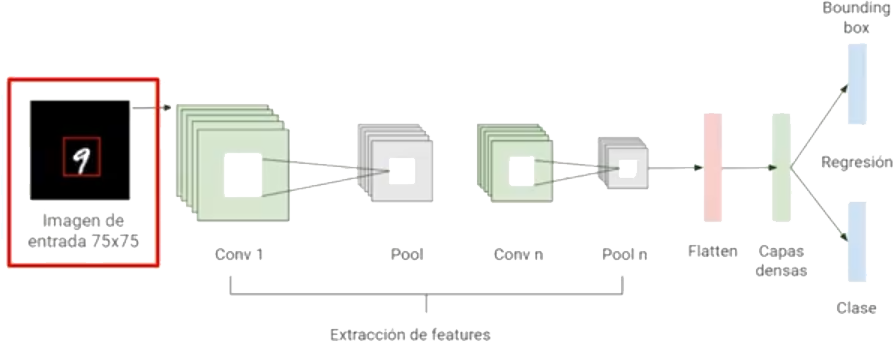
Esta compuesta de dos etapas sucesivas:

* **Localización del objeto:** Define una frontera rectangular llamada *bounding box* alrededor del objeto a detectar.
* **Detección del objeto:** Le asigna una etiqueta al objeto.



**Arquitectura de la detección de objetos**

Se utilizan redes neuronales convolucionales que incluyen las dos etapas —localización y detección.



**Métricas relevantes**

Además de las dos métricas comunes —*accuracy* y *loss*— se tiene otra métrica:

**Intersección entre uniones (*IoU*)**

Se utiliza para medir el solapamiento de dos o más *bounding boxes*. Se define de la siguiente manera.

$$  
\text{IOU} = \frac{\text{Intersection}}{Union}  
$$

Va de 0 a 1 donde 1 significa que las dos regiones son idénticas.

***Object detection and selective search***

El sistema utiliza pequeñas ventanas que recorren toda la imagen.

Existe un problema cuando el objeto es más grande que estas ventanas, pues será categorizado múltiples veces en varios *bounding boxes* que se resuelve con *selección selectiva*,

***Non maximum suppresion* (*NMS*)**

Es una variable que define un umbral de *IoU* que define cuando se deben considerar dos o más *bounding boxes* de un mismo objeto. Si una *bounding box* no supera este umbral, es descartada. De lo contrario es almacenada.

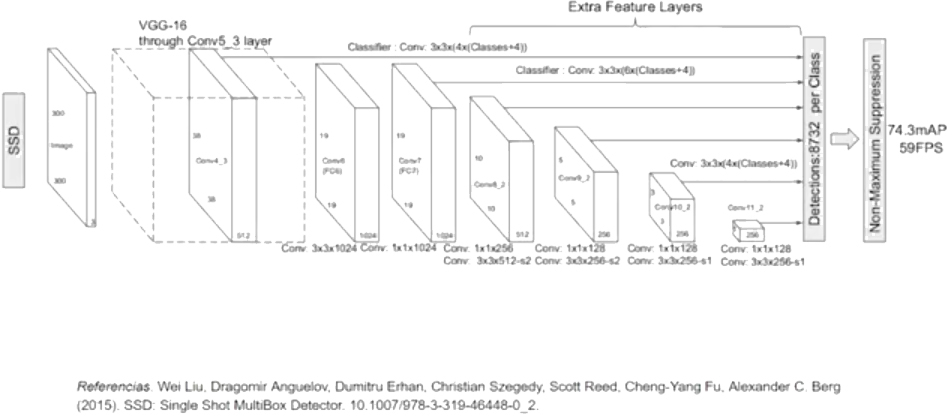
Al final, la *bounding box* elegida es aquella que tiene mayor *IoU*.

**Algortimos más utilizados**

* R-CNN
* Faster R-CNN
* YOLO
* SSD

***Single Shot Detector* (*SSD*)**

Es el algoritmo que se usa en el curso.



***Feature Map***

Así se llama el resultado la detección de *bounding boxes* en *SSD*.

***MobileNet V1***

* Es el clasificador. Es de Google y de bajo coste.
* Aplica la convolución a cada canal por separado —por ejemplo, en una imagen RGB hay un canal R, un canal G y uno B— en lugar de a toda la matriz. Esto ahorra recursos.
* *DeepWise:* Separa las dimensiones de cada color de la matriz en la parte inicial de la red convolucional.
* *PointWise:* Da sentido a los resultados individuales de cada dimensión de color, juntándolos.
* Otra ventaja es que aplican conjuntos de filtros unidos en lugar de los filtros tradicionales.
* Originalmente el clasificaba sólo 1000 objetos.

***MobileNet V2***

* Mejora la *V1* agregando un proceso llamado **expansión**. Si la imagen es pequeña, la expande añadiendo dimensiones hasta 144, haciendo a la red más robusta.
* *Projection Layer* Es una capa que devuelve la imagen a sus dimensiones originales.
* Contempla que existía no linealidad.