



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



TFG del Grado en Ingeniería
Informática

Aplicaciones de Visión
Artificial en Dispositivos de
Edge Computing



Presentado por Miriam Torres Calvo
en Universidad de Burgos — 26 de junio
de 2022

Tutor: Bruno Baruque Zanón



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



D. Bruno Baruque Zanón, profesor del departamento de Ingeniería Informática., área de Ciencia de la Computación e Inteligencia Artificial.

Expone:

Que la alumna D^a. Miriam Torres Calvo, con DNI 45575901K, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado título de TFG.

Y que dicho trabajo ha sido realizado por la alumna bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 26 de junio de 2022

Vº. Bº. del Tutor:

D. Bruno Baruque Zanón

Resumen

La tecnología avanza continuamente y a velocidades que hace unos años eran inexplicables, contamos con su ayuda en diferentes entornos de la sociedad y cada vez en mayor medida. Pero también, contamos con situaciones en las cuáles está no se encuentra tan fácil de acceder, ya que no se cuenta con la facilidad para transportarla y así poder usarla en muchos mas lugares y mas comodamente.

Así surge este proyecto, con la idea de poder usar un modelo de Machine Learning en dispositivos de Edge Computing, como puede ser la Jetson Nano de NVIDIA.

Para su desarrollo, se contara con el lenguaje Python y el modelo escogido para entrenar ha sido YOLO en su cuarta versión.

Descriptores

Deep Learning, Edge Computing, Jetson Nano, YOLO, Python Object Detection, Trasnsfer Learning

Abstract

Technology is advancing continuously and at speeds that were inexplicable a few years ago, we count on your help in different environments of society and to a greater extent. But also, we have situations in which it is not so easy to access, since it is not easy to transport it and thus be able to use it in many more places. and more comfortably.

This is how this project arose, with the idea of being able to use a Machine Learning model in Edge Computing devices, such as the NVIDIA Jetson Nano.

For its development, the Python language will be used and the model chosen for training has been YOLO in its fourth version.

Keywords

Deep Learning, Edge Computing, Jetson Nano, YOLO, Python Object Detection, Trasnsfer Learning

Índice general

Índice general	iii
Índice de figuras	v
Índice de tablas	vi
Introducción	1
1.1. Estructura de la memoria	2
Objetivos del proyecto	5
2.1. Objetivos Software	5
2.2. Objetivos Técnicos	6
Conceptos teóricos	7
3.1. Deep Learning	7
3.2. Edge Computing	8
3.3. Jetson Nano	9
3.4. YOLO	10
3.5. Object Detection	13
Técnicas y herramientas	15
4.1. Metodología	15
4.2. Lenguaje de programación	16
4.3. Algoritmo de detección	16
Aspectos relevantes del desarrollo del proyecto	21
Trabajos relacionados	23

Conclusiones y Líneas de trabajo futuras	25
Bibliografía	27

Índice de figuras

3.1. Red neuronal convolucional	8
3.2. Estructura Edge Computing	9
3.3. Jetson Nano	10
3.4. Explicación YOLO	11
3.5. Representación del contenido etiquetado YOLO	11
3.6. Arquitectura de la red convolucional Darknet	12
3.7. Detección de objetos	13
4.1. Pasos de la metodología Scrum	15
4.2. Pasos de la detección en R-CNN	17
4.3. Arquitectura red Fast-RCNN	18
4.4. Arquitectura red Faster-RCNN	18
4.5. Comparativa resultados Dataset COCO	19

Índice de tablas

3.1. Salida COCO Dataset	12
------------------------------------	----

Introducción

Hoy en día la tecnología ha avanzado tanto, que es muy fácil contar con ella a la hora de realizar ciertas tareas, pero cada vez la demandamos más para poder trabajar codo con codo con ella, es decir, recibir su ayuda de tal forma que les podamos asignar tareas asegurando que tendrán un porcentaje de acierto igual o superior al que tendría si lo realizásemos cualquiera de nosotros.

Pero generalmente, para poder llevar a cabo estas tareas, se necesitan dispositivos con una gran cantidad de compute, ya que necesitaremos entrenarlo con el objeto u objetos a predecir, siendo está la tarea más importante y la que más capacidad de compute va a necesitar y la que más recursos va consumir. Tras su entrenamiento, volveremos a consumir recursos para su detección, de tal forma que necesitaremos un equipo lo suficientemente potente para poder realizar ambas tareas con efectividad y poder obtener buenos resultados.

Debido a esto, el poder entrenar el modelo en un ordenador lo suficientemente potente y seguidamente poder adaptarlo para poder ser utilizado en dispositivos pequeños como puede ser la Jetson Nano de NVIDIA, y que este dispositivo lo ejecute, sacrificando el porcentaje de acierto pero respetando los tiempos de ejecución, puede facilitar a muchos trabajadores y/o investigadores en sus trabajos ya que pueden tener una herramienta funcional en poco espacio y además fácil de transportar para poder usarla en diferentes lugares.

1.1. Estructura de la memoria

La memoria consta de la siguiente estructura:

- **Introducción:** establece el contexto inicial entorno a la idea que se va a desarrollar, además de la estructura del documento y de los materiales que se van a entregar.
- **Objetivos del proyecto:** objetivos que se desean alcanzar durante el desarrollo del proyecto.
- **Conceptos teóricos:** exponer los conceptos que son necesarios disponer para llevar a cabo el proyecto.
- **Técnicas y herramientas:** muestras las técnicas y las herramientas que se han utilizado durante el desarrollo del proyecto.
- **Aspectos relevantes del desarrollo del proyecto:** recopilación de los aspectos más representativos que han tenido lugar durante el desarrollo del proyecto.
- **Trabajos relacionados:** presentación de trabajos que se encuentran relacionados de manera destacable con el desarrollo o el concepto del proyecto.
- **Conclusiones y líneas de trabajo futuras:** descripción de las conclusiones obtenidas durante la realización del proyecto y tras la misma, así como las posibles líneas de mejora.

Además, junto a la presente memoria se incluyen los siguientes anexos relacionados con el desarrollo del modelo de detección y su correspondiente prueba en el dispositivo de Edge Computing:

- **Plan de Proyecto Software:** presentar la planificación temporal llevada a cabo durante el desarrollo del proyecto, así como un estudio de la viabilidad del desarrollo.
- **Especificación de Requisitos:** describir de forma detallada los objetivos generales y los objetivos del proyecto llevado a cabo.
- **Especificación de diseño:** presentar el diseño final del modelo, describiendo el diseño de datos, procedimental y arquitectónico del desarrollo.
- **Documentación técnica de programación:** en este apartado se describen los conocimientos técnicos más relevantes del proyecto, los cuáles son necesarios para poder continuar con el desarrollo.

- **Documentación de usuario:** apartado dirigido al usuario final, dónde se describen los requisitos necesarios en un dispositivo para poder utilizar la herramienta, la instalación de cada uno de ellos, y un manual de usuario, en el que se mostrarán todas las posibles opciones que dispone la herramienta.

Objetivos del proyecto

En este apartado se van a presentar los objetivos que han marcado el proyecto, tanto a nivel software, como técnico.

2.1. Objetivos Software

- Creación del script que permita entrenar el modelo de detección con las clases seleccionadas, obteniendo como resultado el modelo entrenado (al realizarse mediante YOLO, devolverá un fichero .weights)
- Convertir el modelo YOLO (.weights) a un modelo Tensorflow (.pb) para poder trabajar con él.
- Creación de los scripts que permitan la detección de los objetos, ya sea partiendo de una imagen o de vídeo (cargando un vídeo o conectando la webcam), los cuáles contarán con diferentes flags de acción durante la ejecución.
- Creación de un script de preprocesado de cara a la evaluación de varias imágenes etiquetadas, de tal forma que podamos obtenerla información original en un único fichero.
- Medir la calidad de la evaluación del modelo, es decir, calcular el IoU entre las posiciones originales y las predichas por el modelo, pudiendo obtener su mAP, y resultados sobre la predicción (verdaderos positivos, falsos positivos...), además de que por cada imagen se devolverá la imagen con la posición original, la posición predicha y el IoU.
- Mostrar las predicciones en un csv, que muestre el tiempo en el que se ha detectado la predicción, el número de objetos predichos en dicho

instante y el tipo de objeto que es, así como la posición o posiciones en las que se ha encontrado.

- Crear un script, que permita identificar las clases con un tracker, es decir, que identifique las clases y nos las vaya etiquetando según vaya detectando.

2.2. Objetivos Técnicos

- Convertir el modelo a uno apto para la características de la Jetson Nano.
- Usar la plataforma *GitHub* para la organización y gestión del proyecto.
- Seguir los principios de la *metodología ágil Scrum*.
- Usar herramientas que permitan medir la calidad del código.

Conceptos teóricos

Para la compresion de este proyecto, se deben conocer los siguientes conceptos:

3.1. Deep Learning

El Deep Learning [1] es una rama del Machine Learning, donde los algoritmos inspirados en el funcionamiento del cerebro humano (redes neurales) aprenden a partir de grandes cantidades de datos y tratan con un alto número de unidades computacionales.

Gracias a la neurociencia, el estudiio de casos clínicos de daño cerebral sobrevenido y los avances en diagnóstico por imágenes sabemos que hay centros específicos del lenguaje, que existen redes especializadas en detectar diferentes aspectos de la visión, como los bordes, la simetría, áreas relacionadas con el reconocimiento de rostros y las expresiones emocionales de los mismos. Los modelos de Deep Learning imitan estas características de arquitectura del sistema nervioso, permitiendo que dentro del sistema global haya redes de unidades de proceso que se especialicen en la detección de determinadas características que se encuentran ocultas en los datos. Dicho enfoque, ha permitido obtener mejores resultados si los comparamos con la redes monolíticas de neuronas artificiales

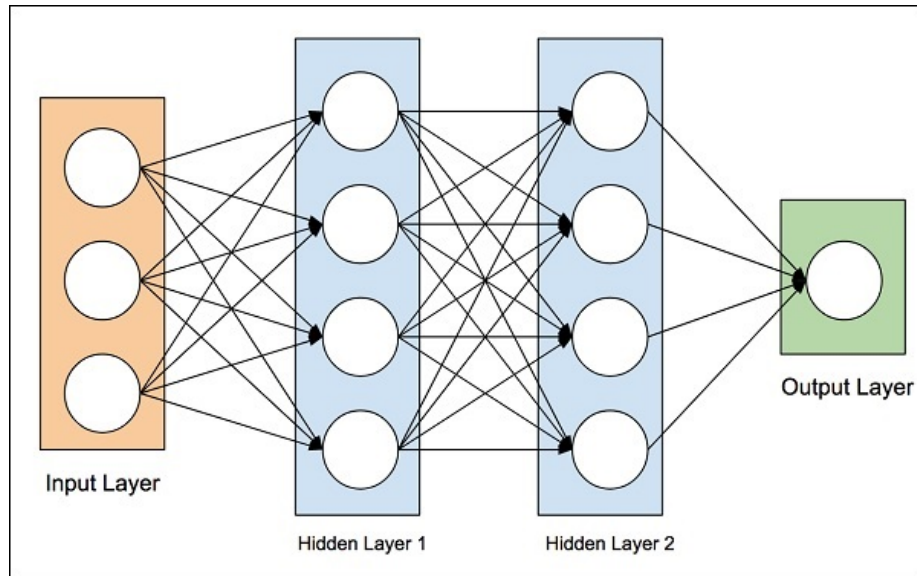


Figura 3.1: Red neuronal convolucional

3.2. Edge Computing

El Edge Computing [2] es un tipo de informática que ocurre, en la ubicación física del usuario, en la ubicación de la fuente de los datos o cerca de estas. Permitiendo que los usuarios obtengan servicios mas rápidos y fiables.

La ventaja fundamental de esto, es que permite a las empresas analizar los datos que sean importantes casi en tiempo real, un hecho que en áreas como la fabricación, la sanidad, las telecomunicaciones o la industria financiera, es una necesidad latente y continua.

Las necesidades industriales hacen que esta tecnología cada vez sea más demandada, debido a que en ciertos entornos la única forma de poder automatizar más los procesos, consiste en tratar de evitar lo máximo posible la comunicación con la nube, consiguiendo reducir las latencias, consumir menos ancho de banda y por su puesto acceder de manera inmediata a análisis y evaluación del estado los sensores y dispositivos que la constituyen.

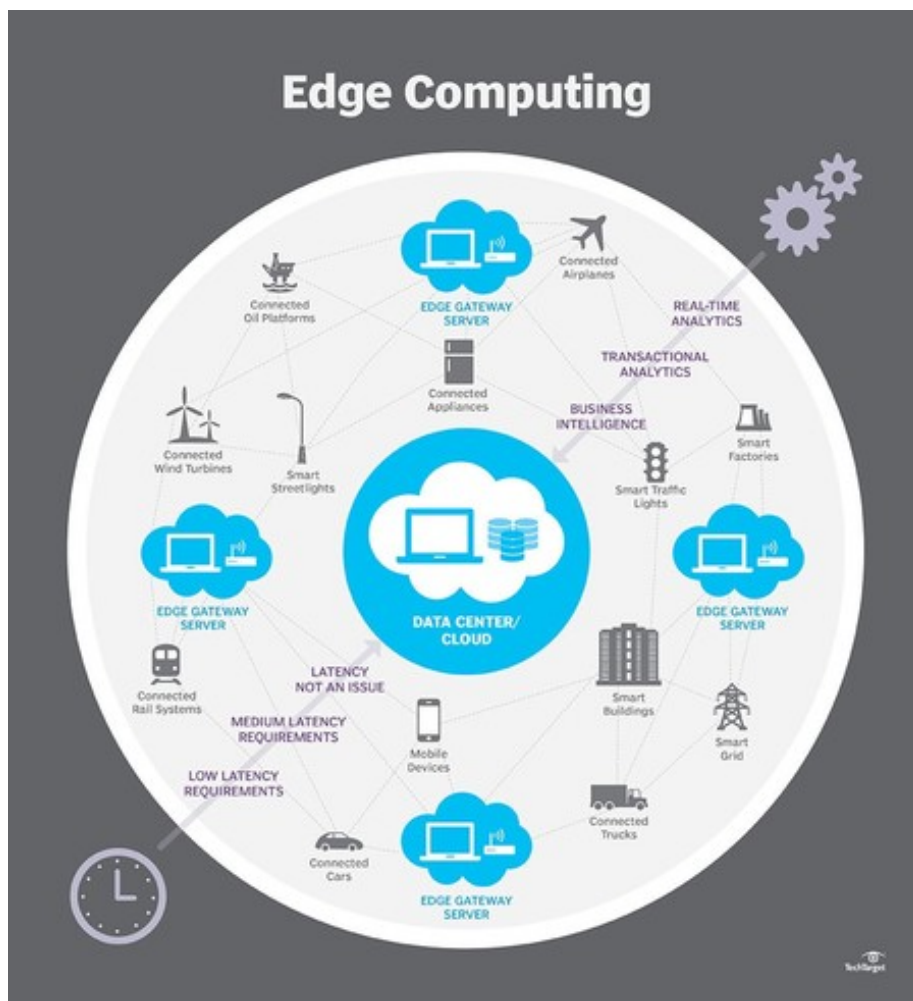


Figura 3.2: Estructura Edge Computing

3.3. Jetson Nano

Una Jetson Nano [3] es un mini PC de bajo coste, el cual cabe en una mano. Se encuentra compuesto por un SoC, procesador ARM de 64 bits de 4 núcleos y una GPU con arquitectura Maxwell con 128 núcleos de procesamiento gráfico, conectividad de red, contando con una potencia total de 472 Gflops, cuenta a su vez, con puertos USB-A, salidas de vídeo HDMI y DisplayPort y un puerto para su conexión a Internet.



Figura 3.3: Jetson Nano

3.4. YOLO

You Only Look Once (YOLO) [4] es un algoritmo de detección que usa Deep Learning y CNN para ello, como su nombre indica sólo necesita mirar la imagen una única vez, de tal forma que la detección es mucho más rápida que en otros algoritmos, pero a cambio de sacrificar rendimiento a la hora de predecir. Para llevar a cabo la detección, divide la imagen en una cuadrícula de $S \times S$ (imagen de la izquierda). Por cada cuadrícula, predice N posibles "bounding boxes" calcula la probabilidad de cada una de ellas, es decir, en total se predicen $S \times S \times N$ cajas diferentes (la gran mayoría con una probabilidad muy baja) (imagen del centro). Por último, se eliminan las cajas que están por debajo de un límite, conocido este como non-max-suppression, de tal forma, que se eliminan los objetos detectados por duplicado, dejando los que poseen un mayor valor de predicción (imagen de la derecha).

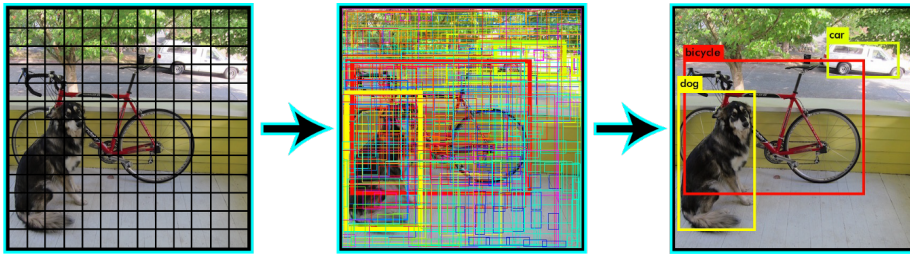


Figura 3.4: Explicación YOLO

Para entrenar un modelo, basado en el algoritmo YOLO, tendremos que tener las imágenes con las que vamos a entrenar etiquetadas con un contenido como el siguiente:

```
0 0.16937475 0.7628560000000001 0.14848899999999998 0.06350499999999999
```

Figura 3.5: Representación del contenido etiquetado YOLO

El primer parámetro representa el id de la clase, es decir, cuando se etiquetan las imágenes se creará un fichero llamado `classes.txt`, con los nombres de todas las clases que se han etiquetado para el modelo. El segundo representa la distancia desde la coordenada 'x' al centro, mientras que el tercero hace lo propio desde la coordenada 'y'. El cuarto parámetro representa el ancho de la anotación, es decir, el ancho del recuadro que conforma la anotación, el último parámetro representa el alto de la anotación.

YOLO utiliza una red CNN llamada Darknet, aunque puede ser utilizada cualquier otra red convolucional a la hora de entrenar. Además, YOLO utiliza las redes convolucionales al final de la cadena, sin necesidad de tener que convertir a una red *tradicional*. La principal crítica que tiene es que a pesar de ser rápida, obtiene peores resultados que las redes *R-CNN*, pero con el paso del tiempo las nuevas versiones que se han lanzado se centran en mejorar dicha precisión en los *bounding boxes*, respetando su eficacia.

La arquitectura de la red se basa en una red convolucional **GoogLeNet** [5], la cuál consta de 24 capas convolucionales. Enbebe en su salida tanto la parte que clasifica la imágenes como la de posicionamiento y tamaño de los objetos.

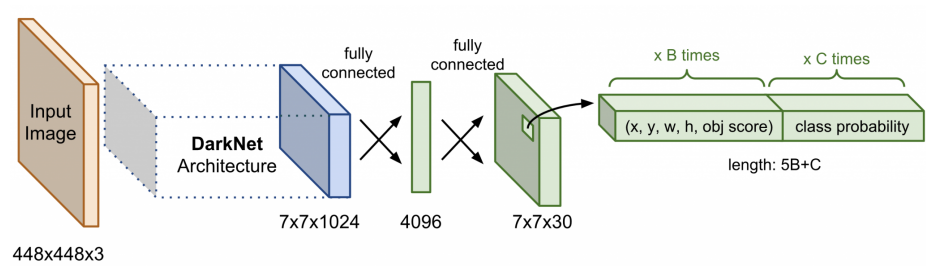


Figura 3.6: Arquitectura de la red convolucional Darknet

Por ejemplo, para el **CocoDataset**, la cuál debe detectar 80 objetos diferentes, nos dará una salida:

Tamaño Grilla	Cantidad Anclas	Cantidad Clases	Ccore, X, Y, Alto, Ancho
13*13	*3*	(80 +	* 5)

Tabla 3.1: Salida COCO Dataset

3.5. Object Detection

El Object Detection [6] es una técnica de visión por ordenador que permite localizar imágenes y/o vídeos. Estos algoritmos se aprovechan del aprendizaje automático o del profundo con el objetivo de obtener resultados significativos, es decir, intentan replicar la inteligencia humana a la hora de reconocer un objeto.

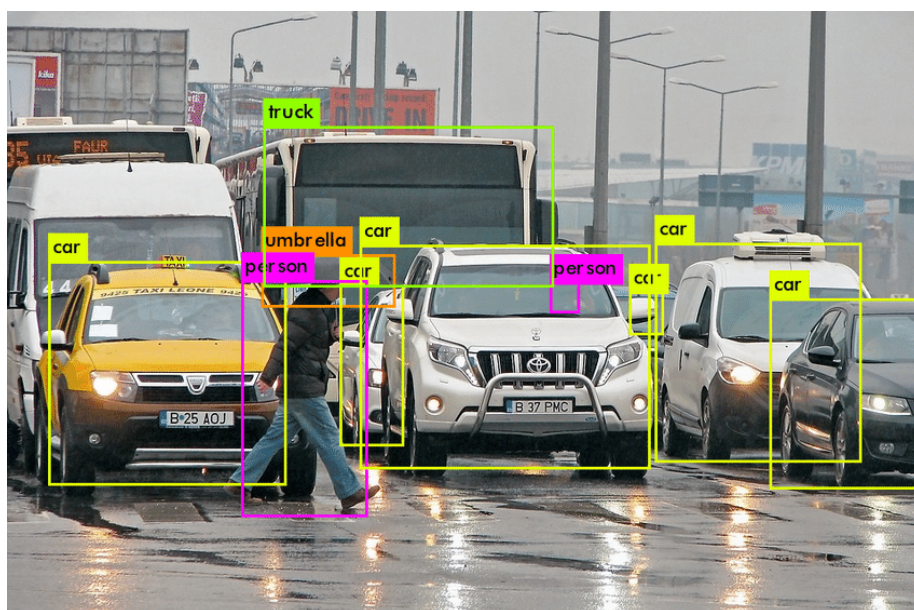


Figura 3.7: Detección de objetos

Técnicas y herramientas

4.1. Metodología

A lo largo del proyecto se intentado seguir la *metodología ágil Scrum*, pero adaptada ya que para poder aplicar esta metodología es necesario contar con un equipo, en el cuál los diferentes miembros se reparten las roles entre los diferentes miembros que lo conforman. En este caso, los roles recaen todo sobre una única persona.

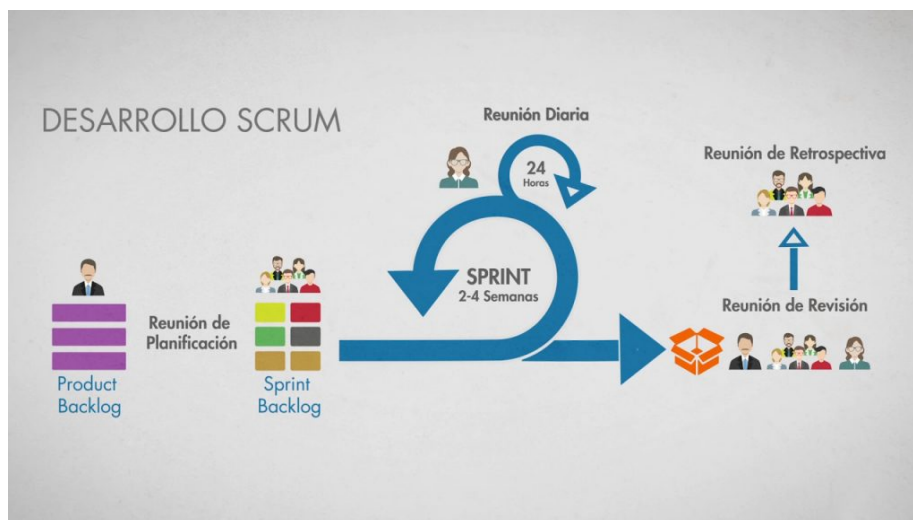


Figura 4.1: Pasos de la metodología Scrum

En primer lugar se encuentra el *Product Backlog* [7] que se trata del alcance del proyecto, el cuál va variando dependiendo de los *feedbacks* que se van obteniendo en cada *sprint*.

Seguidamente, se encuentra el *Sprint Backlog*, dónde se marcan los requerimientos que deben de alcanzar durante el *sprint* que se va a iniciar, es decir, se trata de acortar las tareas de cada uno de los *sprints*.

La siguiente etapa es el *Sprint*, en la cuál tiene lugar la planificación, la implementación, revisión y retrospectiva de la nueva característica software. Esta etapa suele tener una duración de una a dos semanas.

Como último paso del proceso, se encuentra el *incremento del producto*, esta fase consiste en tener una reunión con el cliente con la nueva característica en funcionamiento con el objetivo de obtener una *retroalimentación* por parte del cliente y así volver a empezar el proceso de nuevo.

4.2. Lenguaje de programación

A la hora de empezar un nuevo proyecto es importante relacionado con el *Machine Learning* y el *Edge Computing*, es muy importante seleccionar el lenguaje, con el cuál queremos trabajar destacando dos: **Python** [8] y **Matlab** [9]. En este caso se decantó por el uso de *Python*, debido al mayor conocimiento de este lenguaje y haber trabajado más con este lenguaje que con *Matlab*. No hay grandes ventajas entre escoger uno u otro.

4.3. Algoritmo de detección

A parte de tener claro el lenguaje que se quiere utilizar, otra característica a tener en cuenta es elegir el *algoritmo de detección* en el que se va a basar el modelo. Existen diferentes algoritmos:

- **CNN:** (*Convolutional Neuronal Network*) es la opción más básica que se puede escoger, ya que se parte de una red neuronal convolucional [10] la cuál itera la imagen hasta devolver las posiciones de los objetos que detecta.

Esta opción trae consigo diferentes inconvenientes:

- Si la imagen detecta varios objetos, situados en zonas opuestas, ¿cuántos píxeles tendremos que desplazarnos en cada dirección?.
- El tiempo de cómputo es variable, pudiendo llegar a ser muy largo, ya que por cada movimiento implica una clasificación individual con la red.
- Detectar un objeto dentro de la red, no indica que se poseen los valores 'x' e 'y' de su posición.

- Si por un casual el desplazamiento de píxeles que se realiza es muy pequeño, podríamos estar detectando el mismo objeto múltiples veces.
 - Si dos objetos se encuentran muy juntos, se podrían llegar a detectar como un único objeto.
- **R-CNN:** (*Region Based Convolutional Neural Networks*) surgen en el año 2014, con la siguiente propuesta: determinar primero las regiones de interés de la imagen y después realizar la clasificación de imágenes sobre dichas áreas usando una red preentrenada.^[11] Esto implica, que haya un primer algoritmo que detecte las áreas de interés de la imagen, las cuáles pueden ser muchas y de diversos tamaños. Seguidamente, se pasan las diferentes regiones por la CNN, validándose las clases correctas mediante un clasificador binario, de tal forma, que se eliminarán las que tenga un bajo nivel de confianza. Por último, se ajustaría la posición mediante un regresor.

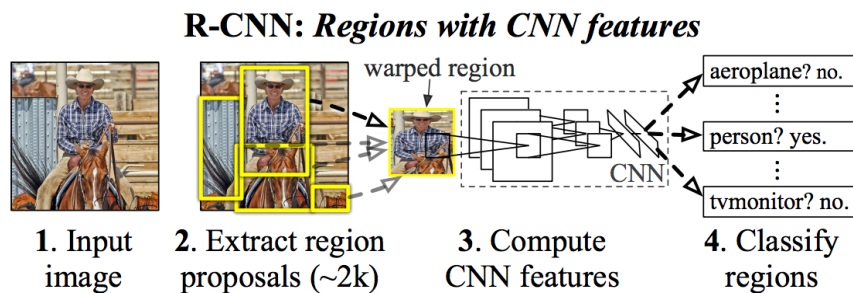


Figura 4.2: Pasos de la detección en R-CNN

- **Fast R-CNN / Faster R-CNN:** Son dos algoritmos que surgen como mejora a R-CNN: [12]
- **Fast R-CNN:** mejora el algoritmo inicial reutilizando algunos recursos, como las *features* extraídas por la CNN, de tal forma que se agiliza el entreno y la detección de las imágenes. Esta red, posee también mejoras en el cálculo del IoU (*Intersection Over Union*) y en la función de *Loss*. Pero a pesar de esto, no tiene mejoras drásticas en la velocidad de entrenamiento y en la detección.

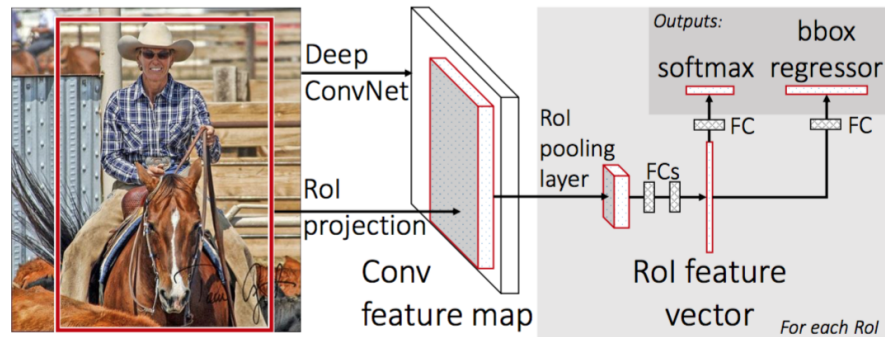


Figura 4.3: Arquitectura red Fast-RCNN

- **Faster R-CNN:** logra una mejora de velocidad al integrar el algoritmo de *region proposal* [13] sobre la propia CNN. Además aparece el concepto de usar *anchors* fijos, lo cuál consiste en usar tamaños pre calculados para la detección de objetos de la red.

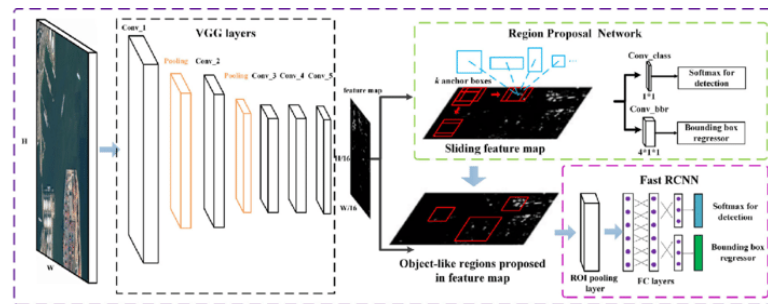
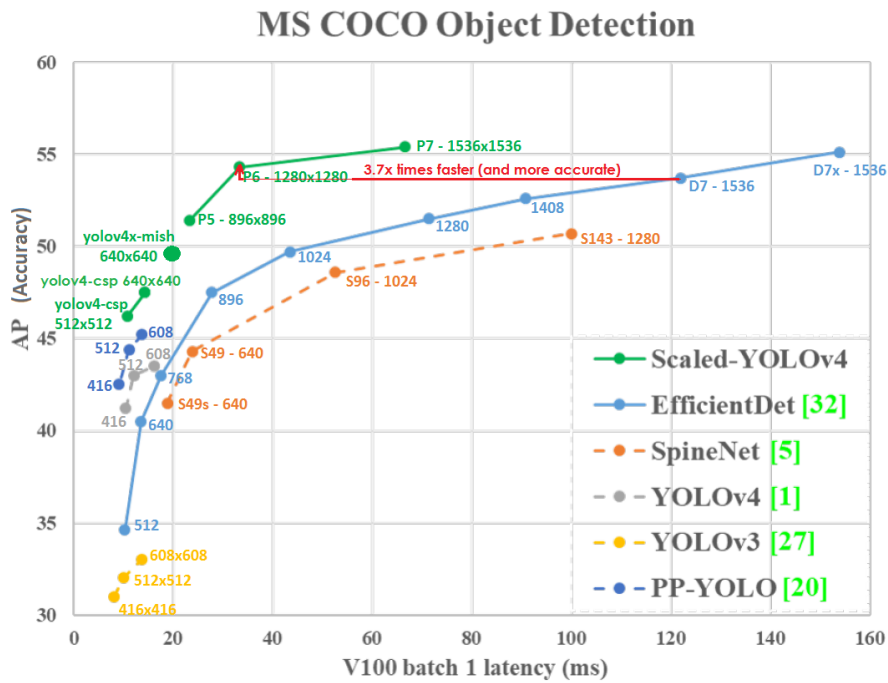


Figura 4.4: Arquitectura red Faster-RCNN

- **YOLO:** surge en 2016, su nombre viene formado por las siglas de *You Only Look Once*. [4] Esta red, como su propio nombre indica hace una única pasada a la red y detecta todos los objetos para los que ha sido entrenada para clasificar, al realizar un único vistazo obtiene velocidades muy buenas en equipos que no son necesariamente potentes. Lo cuál permite, detecciones en tiempo real de cientos de objetos de forma simultánea y su ejecución en dispositivos móviles.

Debido a esto, el modelo escogido ha sido YOLO y en su versión 4, la cuál fue lanzada en Abril del año 2020.



Aspectos relevantes del desarrollo del proyecto

Este apartado pretende recoger los aspectos más interesantes del desarrollo del proyecto, comentados por los autores del mismo. Debe incluir desde la exposición del ciclo de vida utilizado, hasta los detalles de mayor relevancia de las fases de análisis, diseño e implementación. Se busca que no sea una mera operación de copiar y pegar diagramas y extractos del código fuente, sino que realmente se justifiquen los caminos de solución que se han tomado, especialmente aquellos que no sean triviales. Puede ser el lugar más adecuado para documentar los aspectos más interesantes del diseño y de la implementación, con un mayor hincapié en aspectos tales como el tipo de arquitectura elegido, los índices de las tablas de la base de datos, normalización y desnormalización, distribución en ficheros³, reglas de negocio dentro de las bases de datos (EDVHV GH GDWRV DFWLYDV), aspectos de desarrollo relacionados con el WWW... Este apartado, debe convertirse en el resumen de la experiencia práctica del proyecto, y por sí mismo justifica que la memoria se convierta en un documento útil, fuente de referencia para los autores, los tutores y futuros alumnos.

Trabajos relacionados

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final grado no parece obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.

Conclusiones y Líneas de trabajo futuras

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.

Bibliografía

- [1] IBM, “Qué es deep learning.” [Online]. Available: <https://www.ibm.com/es-es/cloud/watson-studio/deep-learning>
- [2] R. Hat, “Qué es edge computing,” 2021. [Online]. Available: <https://www.redhat.com/es/topics/edge-computing/what-is-edge-computing>
- [3] M. Computer, “Nvidia jetson nano, una raspberry pi para ia,” 2019. [Online]. Available: <https://www.muycomputer.com/2019/03/19/nvidia-jetson-nano/>
- [4] E. A., “Detección de objetos con yolo: implementaciones y como usarlas.” [Online]. Available: <https://medium.com/@enriqueav/detecci%C3%B3n-de-objetos-con-yolo-implementaciones-y-como-usarlas-c73ca2489246>
- [5] R. Alake, “Deep learning: Googlenet explained,” 2020. [Online]. Available: <https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765>
- [6] Mathworks, “What is object detection?” [Online]. Available: <https://es.mathworks.com/discovery/object-detection.html>
- [7] Proaglist, 2016. [Online]. Available: <https://proagilist.es/blog/agilidad-y-gestion-agil/agile-scrum/los-11-pasos-para-implementar-metodologia-scrum/>
- [8] Oracle, “¿qué es python?” [Online]. Available: <https://developer.oracle.com/es/python/what-is-python/>
- [9] MathWorks, “Matlab.” [Online]. Available: <https://es.mathworks.com/products/matlab.html>

- [10] Nebulova, “Predicción con redes neuronales convolucionales (cnn),” 2021. [Online]. Available: <https://www.nebulova.es/blog/redes-neuronales-convolucionales>
- [11] S. Neelam, “Introduction to object detection with rcnn family models,” 2021. [Online]. Available: <https://medium.com/analytics-vidhya/introduction-to-object-detection-with-rcnn-family-models-310558ce2033>
- [12] R. Gandhi, “R-cnn, fast r-cnn, faster r-cnn, yolo — object detection algorithms,” 2018. [Online]. Available: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>
- [13] T. Karmarkar, “Region proposal network (rpn) — backbone of faster r-cnn,” 2018. [Online]. Available: <https://medium.com/egen/region-proposal-network-rpn-backbone-of-faster-r-cnn-4a744a38d7f9>