



CARRERA DE ESPECIALIZACIÓN EN INTELIGENCIA ARTIFICIAL

MEMORIA DEL TRABAJO FINAL

Reconocimiento de intrusos en video de vigilancia aérea

Autor:

Ing. Juan Pablo Nieto Uribe

Director:

Esp. Ing. Hernán Contigiani (FIUBA)

Jurados:

Nombre del jurado 1 (pertenencia)

Nombre del jurado 2 (pertenencia)

Nombre del jurado 3 (pertenencia)

*Este trabajo fue realizado en la ciudad de Bogotá, Colombia,
entre octubre de 2021 y junio de 2023.*

Resumen

La presente memoria detalla el proceso de creación de un módulo de inteligencia artificial, diseñado para implementar un sistema de detección de personas y animales de gran porte de forma que sea posible automatizar el proceso de vigilancia y así reducir el personal encargado a esta labor, y permitiendo abaratar costos y aumentar la precisión en el proceso. En particular, lo que se pretende con el desarrollo de este trabajo es detectar intrusos en *streaming* de vídeo tomado desde un dron, como parte de un sistema más grande de vigilancia, que será implementado por la empresa Manta Beach. Esta empresa, que hace parte del programa de vinculación de la universidad, propuso el trabajo como parte de su plan de investigación y desarrollo.

Para la creación de este módulo, se hace un uso extensivo de los conocimientos de la rama de visión por computadora a través de redes neuronales convolucionales. De igual manera, se ha requerido una base sólida en Python, teniendo en cuenta que se partió de un algoritmo establecido con anterioridad.

Agradecimientos

Índice general

Resumen	I
1. Introducción general	1
1.1. Contexto de la problemática	1
1.2. Motivación	1
1.3. Estado del arte	2
1.4. Objetivos y alcance	4
2. Introducción específica	7
2.1. Definición de intruso	7
2.2. Modelo utilizado	8
2.3. Protocolos de vídeo y transmisión	10
2.4. Software y hardware utilizados	11
3. Diseño e implementación	13
3.1. Consideraciones generales	13
3.2. Arquitectura del proyecto	14
3.3. Esquema del módulo	14
3.4. Esquema de modificaciones	14
3.5. Ajustes para cumplir con el rendimiento	14
3.6. Entrega de resultados al resto del software	14
4. Ensayos y resultados	15
4.1. Descripción del proceso de pruebas	15
4.2. Pruebas en Raspberry Pi	15
4.3. Caso de Uso	15
5. Conclusiones	17
5.1. Resultados Obtenidos	17
5.2. Tiempos de ejecución	17
5.3. Monitoreo de resultados	17
A. Tablas extendidas	19
Bibliografía	21

Índice de figuras

1.1. Esquema de redes neuronales de YOLOv3. ¹	4
--	---

Índice de tablas

1.1. Fecha de lanzamiento de distintos modelos YOLO.	3
2.1. Clases detectadas como intruso por el módulo.	8
2.2. Software requerido para el desarrollo y la ejecución del módulo. . .	11
A.1. Software requerido para el desarrollo y la ejecución del módulo . .	19

A Iván. Por tu intuición.

Capítulo 1

Introducción general

En este capítulo se expone la problemática, su contexto y el portafolio de opciones que se encuentran disponibles para darle solución. De igual manera, se hace una descripción rápida de cómo se acotó el problema a fin de poder generar un mínimo producto viable.

1.1. Contexto de la problemática

El uso de cámaras de vídeo es una práctica extremadamente común en el negocio de la seguridad. Su uso puede mejorar el área de cobertura de vigilancia [1], mientras que al mismo tiempo reduce la cantidad de personal empleado en hacer rondas de seguridad. Sin embargo, tener cámaras estáticas implica que será necesario contar con más de ellas [2], que a la vez tendrán que ser revisadas por varias personas.

La propuesta abordada en este trabajo busca solucionar estos dos problemas a través de la creación de un módulo de inteligencia artificial que permita el análisis automático de varios *streamings* de vídeo provenientes de cámaras montadas en drones, y sea capaz de responder a estos estímulos digitales, al pasarle la información del intruso encontrado al software del que hará parte.

Esto permitirá que haya un análisis mucho más preciso y a menor costo de lo que era posible anteriormente, teniendo en cuenta que los drones podrán abarcar un área mucho mayor, y que no se dependerá de la concentración de una persona para encontrar intrusos. Esto implica que el sistema deberá estar en capacidad de analizar más de un *streaming* de video al tiempo. Igualmente, el uso de estas tecnologías permite la utilización de cámaras infrarrojas, lo que hace posible la identificación de intrusos en la noche, cuando el ojo humano es incapaz de distinguir.

Dado que este sistema hará parte de un módulo más grande, se decidió que, dentro de los requisitos, el sistema no deberá tomar decisiones basadas en sus hallazgos, sino simplemente comunicarlos tanto a otro sistema (del que hace parte), como al personal de seguridad encargado.

1.2. Motivación

Este trabajo responde a cinco necesidades concretas del ámbito de la vigilancia. Estas necesidades son:

- Razones económicas: teniendo en cuenta que, al reducir al personal encargado de hacer seguimiento a las cámaras, o de hacer rondas de seguridad en horarios determinados, es posible reducir costos. De igual manera, el uso de drones permite una reducción sustancial en el número de cámaras que deberán ser instaladas, cosa que resulta particularmente útil en grandes áreas y permitirá una reducción muy importante en los costos de mantenimiento asociados.

Al ser tan monótono, es muy probable que el personal encargado se vea expuesto a distracciones, que pueden tener consecuencias severas en el momento en el que se presente algún evento que comprometa la seguridad. La situación se puede agravar en caso de tener una gran cantidad de cámaras, dado que el estar constantemente obligados a supervisarlas puede generar niveles no razonables de carga en el personal.

- Velocidad de reacción: una vez identificada una potencial amenaza, a través del sistema encargado de controlar los drones se podrá enviar órdenes a los equipos a fin de tomar mejores decisiones para manejar la situación. Es por esto que es de vital importancia que el módulo de inteligencia artificial encargado de la detección de los intrusos también sea capaz de hacerlo en tiempos razonables, que en este caso, han sido definidos por el cliente y se encuentran por debajo de los 4 segundos. Este valor se ha calculado a partir de la velocidad esperada de vuelo de los drones de hasta 30 km/h, lo que otorga un radio máximo de 33 metros entre detecciones. Cabe anotar, sin embargo, que el tiempo ideal es menor o igual a un segundo, lo que genera un radio de 8 metros aproximadamente.
- Aporte de pruebas: los drones (a través de la necesidad de velocidad de reacción) son un método ideal para aportar pruebas confiables de los hechos. Es así como un modelo de inteligencia artificial permite generar una mayor claridad en la interpretación de las pruebas aportadas, lo que puede generar una mejor comprensión y análisis de los hechos por parte de todos los implicados.
- Identificación de intrusos en horas de la noche: teniendo en cuenta que los modelos de inteligencia artificial pueden ser capaces de identificar las clases requeridas con diferentes condiciones de iluminación, puede ayudar a la identificación de intrusos cuando las condiciones de iluminación no son ideales. Por otro lado, es muy probable que esto sea mejorado sustancialmente con el uso de cámaras infrarrojas.

Este tipo de propuestas, sin embargo, no se han hecho posibles hasta hace relativamente poco, cuando los modelos de inteligencia artificial alcanzaron una madurez que permite que no requieran grandes capacidades computacionales para poder ser ejecutados, de forma que implementarlos no requiera una inversión importante en equipos de cómputo.

1.3. Estado del arte

Existen varias técnicas y varios modelos que se pueden utilizar para implementar visión por computadora. Uno de los métodos más comunes a la fecha es el algoritmo YOLO, que se caracteriza por ser el primer modelo que no requiere hacer

TABLA 1.1. Fecha de lanzamiento de distintos modelos YOLO.

Versión	Fecha de lanzamiento
YOLOv3	Abril de 2018
YOLOv3 Lite	Abril de 2020
YOLOv6	Junio de 2022

un doble análisis de la imagen [3] (de aquí proviene su nombre: *You Only Look Once*), y en consecuencia, analiza las imágenes en un tiempo mucho menor.

A la fecha de redacción de este documento, las versiones más avanzadas de los modelos de la familia YOLO son las versiones 6 y 7, que ofrecen grandes mejoras en términos de precisión y rendimiento en comparación con sus antecesores. Sin embargo, como se puede ver en la tabla 1.1, estos modelos no estuvieron disponibles hasta una fecha posterior al planteamiento del trabajo.

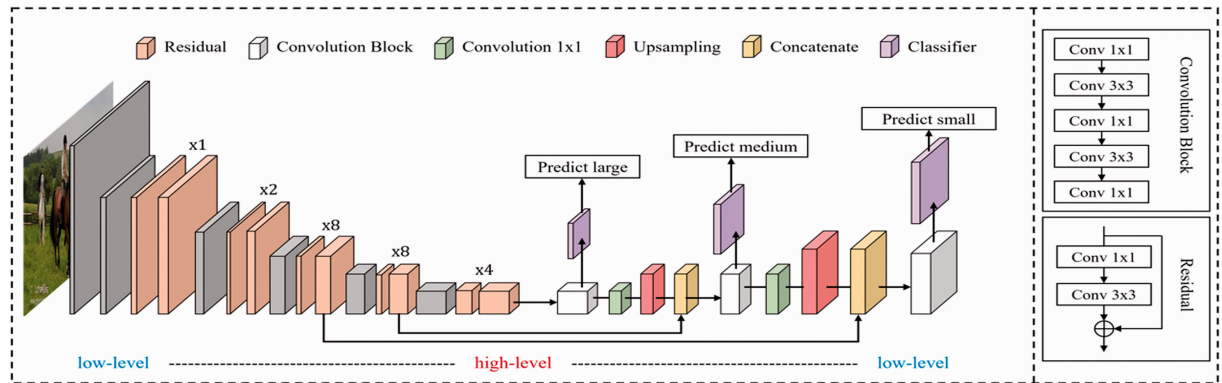
Por otro lado, las versiones 4 y 5 fueron descartadas para la implementación de este trabajo, visto que son dos versiones que fueron desarrolladas por personas diferentes a los desarrolladores originales, quienes se retiraron del proyecto por conflictos éticos con las capacidades de la técnica y sus potenciales usos [4]. En consecuencia, es muy extendido el uso de la versión 3 del modelo a pesar de no ser el más actualizado.

Ahora bien, es importante tener en cuenta que existen diferentes versiones del modelo, cada una adaptada a algún caso de uso particular. Destacan dentro de estas versiones la arquitectura YOLOv3 y la arquitectura YOLOv3 Lite:

- YOLOv3: es la versión “base” de la arquitectura YOLO. Representa un avance incremental con respecto a las arquitecturas YOLOv1 y YOLOv2. Su principal ventaja por encima de los modelos anteriores (en particular comparado con RetinaNet-50 y RetinaNet-101) es una reducción significativa en los tiempos de detección. El *dataset* sobre el que fue entrenado es el *dataset* COCO.
- YOLOv3 Lite: es una versión reducida de la versión YOLOv3 diseñada específicamente para ser ejecutada en dispositivos con menor potencia, o que no cuentan con tarjeta gráfica [5]. Si bien es un modelo que es capaz de reducir sustancialmente la complejidad del equipo requerido, también cuenta con una menor precisión en el momento de hacer la detección correspondiente.

Por otro lado, además de la arquitectura, es de vital importancia tener en cuenta el *dataset* que fue utilizado para el entrenamiento del modelo. En este caso, se trata del *dataset* COCO, que cuenta con más de 200.000 imágenes etiquetadas (a la fecha de redacción de este documento), con más de 1.5 millones de objetos detectados en ellas. [6]

A continuación se puede ver una ilustración del funcionamiento de las redes neuronales convolucionales a través de las que funciona el modelo YOLOv3:

FIGURA 1.1. Esquema de redes neuronales de YOLOv3.¹

1.4. Objetivos y alcance

A continuación, se presentan los requisitos establecidos en la etapa de apertura del trabajo y su estado actual a la fecha. Es necesario tener en cuenta que, debido a cambios operativos, algunos de estos requisitos no pudieron ser cumplidos en su totalidad.

1. Requerimientos funcionales:

1.1 El módulo debe detectar personas o animales de gran porte en áreas que el cliente defina como restringidas.

- Tipo: obligatorio
- Estado: finalizado

1.2 El módulo debe ser capaz de detectar a un intruso en un tiempo no mayor a 4 segundos.

- Tipo: obligatorio
- Estado: finalizado

1.3 El módulo debe correr en un computador de uso de hogar. Este debe ser comparable o menor a un Intel Core I5 y con no más de 8 GB de memoria RAM.

- Tipo: obligatorio
- Estado: finalizado

1.4 El sistema debe ser capaz de reconocer la presencia de intrusos, aun cuando el vídeo se haya registrado con el dron en movimiento.

- Tipo: obligatorio
- Estado: finalizado

1.5 **Requerimiento Opcional:** el sistema debe correr en un equipo que quepa a bordo de los drones (una Raspberry Pi, por ejemplo).

¹Imagen tomada de https://journals.sagepub.com/cms/10.1177/0020720920983524/asset/images/large/10.1177_0020720920983524-fig2.jpeg

- Tipo: opcional
- Estado: descartado
- 2. Requerimientos de documentación:
 - 2.1 El sistema deberá contar con un manual de usuario.
 - Tipo: obligatorio
 - Estado: finalizado
 - 2.2 Se deberá contar con un manual de instalación del módulo desarrollado.
 - Tipo: obligatorio
 - Estado: finalizado
 - 2.3 El código fuente deberá estar debidamente documentado.
 - Tipo: obligatorio
 - Estado : finalizado
- 3. Requerimientos de pruebas:
 - 3.1 El módulo deberá ser probado formalmente según indicaciones del cliente.
 - Tipo: requisito modificado
 - Estado: descartado
 - 3.2 El cliente debe poder acceder en cualquier momento a la versión más reciente del código a través de GitHub.
 - Tipo: obligatorio
 - Estado: finalizado
- 4. Requerimientos de la interfaz:
 - 4.1 El usuario debe poder iniciar fácilmente la ejecución del módulo.
 - Tipo: obligatorio
 - Estado: finalizado
- 5. Requerimientos de interoperabilidad:
 - 5.1 **Opción A:** el módulo debe estar en capacidad de recibir el *streaming* de vídeo de entre 4 y 8 drones.
 - Tipo: requisito modificado
 - Estado: descartado
 - 5.2 **Opción B:** el módulo debe ser capaz de recibir un *streaming* de vídeo y ejecutar la detección de intrusos en hardware dedicado a bordo.
 - Tipo: requisito modificado
 - Estado: descartado

5.3 Se requiere que el sistema pueda conectarse con otro módulo capaz de tomar decisiones con base en la información de detección de intrusos.

- Tipo: obligatorio
- Estado: finalizado

A pesar de que algunos de los requerimientos funcionales fueron descartados dados algunos cambios que se presentaron en las plataformas del cliente, también hubo requerimientos que fueron agregados:

6. Requerimientos nuevos:

6.1 El módulo deberá retransmitir los *frames* procesados a un servidor indicado por el cliente.

- Tipo: obligatorio
- Estado: finalizado

Es importante tener en cuenta que el trabajo no contempló:

- Reconocimiento del intruso: es decir, en caso de que se detecte a una persona, no se deberá dar con su identidad.
- Reacción del dron: es decir, una vez se haya detectado satisfactoriamente a un intruso, el sistema no deberá dar instrucciones específicas al dron, por ejemplo, de seguirlo.
- No existen requisitos específicos de cómo realizar la retransmisión de los *frames* ya procesados.

Capítulo 2

Introducción específica

En este capítulo, se presentan los conceptos necesarios para poner en funcionamiento el trabajo de detección. Se inicia con la descripción de los objetos a detectar, seguido de los modelos de visión por computadora disponibles. De igual manera, se abordan los protocolos de transmisión y recepción del vídeo y se finaliza con los requerimientos computacionales necesarios para llevar a cabo la detección.

2.1. Definición de intruso

Para cumplir con el objetivo específico del trabajo, es importante tener en cuenta la definición de intruso dada por el cliente, que incluye la detección de personas y animales de gran porte, lo que implica que la respuesta del módulo deberá activarse con cualquier animal de mayor tamaño al de un perro o un gato. Es importante tener en mente que el *dataset* COCO permite el reconocimiento de una serie de animales de gran porte que no hacen parte de la fauna argentina, por lo que se decidió en conjunto con el cliente que algunas etiquetas se filtrarían a fin de evitar que el modelo intente reconocer estas clases. Algunos ejemplos de etiquetas filtradas incluyen:

- Elefantes
- Jirafas
- Cebras

Por otro lado, y a fin de mejorar el reconocimiento de intrusos, se decidió incorporar algunas clases que corresponden a objetos que suelen cargar las personas y vehículos. De esta manera se hace posible una suerte de reconocimiento indirecto, dándole al modelo una “segunda oportunidad” en casos en los que por algún motivo generó un falso negativo en el reconocimiento de una persona. Es importante tener en cuenta que estos objetos propuestos ya hacían parte del *dataset* COCO, por lo que no se requirió reentrenar el modelo para lograr su detección. Dentro de estos objetos relacionados encontramos:

- Vehículos
- Objetos de porte personal, como maletines o sombrillas

Es importante tener en cuenta que el módulo marcará toda detección como una intrusión. Esto quiere decir que las detecciones en lotes vecinos no se filtrarán dado que se considera que la totalidad del *frame* es área vigilada. Dicho esto, las clases que generarán un reporte positivo para intruso al ser detectadas son:

TABLA 2.1. Clases detectadas como intruso por el módulo.

Clases detectadas como intruso	
Persona	Vaca
Camión	Bus
Automóvil	Sombrilla
Motocicleta	Mochila
Bicicleta	Maletín
Perro	Maleta
Caballo	Gato

Hay que recordar que la sola detección de una de estas clases generará un reporte positivo de intruso. El modelo no está en capacidad de identificar personas específicas en el vídeo, por lo que le resultará imposible diferenciar entre personal autorizado en un área (el mismo personal de seguridad, o la persona al mando del dron), y a intrusos en la zona.

2.2. Modelo utilizado

El modelo utilizado para el desarrollo de la totalidad del trabajo es YOLOv3. La principal ventaja que tiene este modelo es que es capaz de ejecutarse con una velocidad considerablemente mayor a la de modelos anteriores como RetinaNet-50 y RetinaNet-101.

Este modelo se basa en el uso de redes neuronales convolucionales (CNN) que, al dividir la imagen en una cuadrícula, permiten calcular la probabilidad de ocurrencia de objetos en cada celda. Con esta información, el modelo determina tres parámetros clave para su funcionamiento:

- *Intersection over Union* (IoU) [8], se trata del criterio utilizado para eliminar cajas redundantes, y asegurar que solo se mantenga una por cada objeto detectado, incluso si hay varios objetos cercanos en la imagen. De esta manera, se evita la detección de objetos duplicados y se obtiene una detección más precisa y clara de los objetos presentes en la imagen.

Se calcula a través de la fórmula 2.1:

$$IoU = \frac{A_s}{A_u} \quad (2.1)$$

Donde A_s representa el área de superposición entre ambas cajas y A_u el área de unión.

Para lograr esto, cada celda es responsable de calcular la probabilidad de que haya un objeto dentro de sus fronteras, se genera una serie de cajas propuestas alrededor del objeto, y a través del uso de redes neuronales convolucionales se calcula la probabilidad de ocurrencia.

Finalmente, el modelo elimina las cajas que no cumplen con dos criterios importantes: en primer lugar, se descartan aquellas cajas que tienen una probabilidad de ocurrencia menor que la máxima del objeto de la zona. Este algoritmo es conocido como *Non-Maximum Suppression* [9] y utiliza la probabilidad de ocurrencia, también conocida como *objectness score* [10], para

seleccionar solo las cajas más relevantes. En segundo lugar, se eliminan las cajas que tienen un *Intersection Over Union (IoU)* mayor a un valor determinado, que suele ser 0,45 aunque también se puede ajustar como parámetro en el modelo.

- *Offset*: el *offset* se refiere a la distancia medida desde la esquina superior izquierda de la celda con mayor probabilidad de contener un objeto [11]. Esta medida es una forma eficiente de representar la posición de las cajas y reduce significativamente la cantidad de cálculos necesarios, lo que aumenta la eficiencia computacional del modelo [12]. Una vez obtenido el *offset*, se utilizan estos valores para calcular las coordenadas de las cajas que no fueron descartadas.
- *Tensor*: hace referencia a un vector utilizado para representar la probabilidad [13] de que cada uno de los objetos incluidos en el *dataset* COCO [6] se encuentre en la caja. Esta representación no solo permite determinar la posición de diferentes objetos en la imagen, sino también discriminarlos por clases. La clase con mayor probabilidad es entonces reportada por el modelo.

Es importante tener en cuenta que estos no son los parámetros finales arrojados por el modelo, sino que más adelante, la red calcula las coordenadas de la caja con la que se rodeará al objeto. Es decir, para cada objeto detectado, el modelo va a devolver (en ese orden):

- *pc*: *objectness score*, la probabilidad de que haya un objeto.
- *bx*: coordenada x del centro de la caja propuesta.
- *by*: coordenada y del centro de la caja propuesta.
- *bh*: altura de la caja propuesta.
- *bw*: ancho de la caja propuesta.
- *c*: tensor con la probabilidad de ocurrencia de las 80 clases que forman parte del *dataset* COCO.

Esto se logra a través del uso de *Anchor Boxes*: cajas de predicción predefinidas que el modelo va refinando a medida que analiza la imagen y calcula las métricas de cada uno de los objetos encontrados. A través de este método es posible eliminar la necesidad de una “ventana deslizante”, lo que a su vez permite analizar la totalidad de la imagen en un tiempo mucho menor que los métodos basados en aquel método, y, en consecuencia, se hace posible llevar a cabo un análisis en tiempo real.

Dadas las bondades explicadas anteriormente, en conjunto con su facilidad de implementación en Python, se escogió partir del modelo YoloV3 y modificar su código fuente para lograr generar y extraer la información pertinente para completar el propósito del trabajo. Cabe aclarar que la decisión de utilizar este modelo se vio reforzada una vez se ejecutó la primera versión en tanto que el modelo está optimizado para funcionar con la tarjeta gráfica (NVIDIA Titan X), el código se pudo ejecutar satisfactoriamente, sin dejar a un lado los requisitos del trabajo en CPU (AMD Ryzen 5).

2.3. Protocolos de vídeo y transmisión

El módulo del trabajo encargado de extraer y permitir el cálculo sobre cada uno de los *frames* es OpenCV, por lo que los protocolos de vídeo aceptados corresponderán a los aceptados por OpenCV. En este sentido, el código está en capacidad de detectar intrusos en:

- Vídeo guardado localmente
- Secuencia de imágenes
- URL de *streaming* de vídeo
- *Pipeline* Gstreamer
- Datos obtenidos desde *webcam*

Sin embargo, es importante mencionar que, durante el desarrollo, se hicieron pruebas únicamente con los métodos de vídeo guardado localmente y URL de *streaming* de vídeo. Los otros métodos no han sido probados, y, por lo tanto, no se podrá garantizar su correcto funcionamiento al solicitar al módulo ejecutarse utilizándolos como origen de datos.

Por otro lado, y recordando que se trata de la biblioteca que hace posible extraer la información del vídeo para hacer posible su análisis, es de vital importancia tener en mente cuáles son los formatos de vídeo admitidos por OpenCV. Dentro de ellos se encuentran:

- H264
- MPEG4
- AV1

Ahora bien, para acceder a ellos, bastará con pasar el link apropiado al método encargado. Dentro de los protocolos probados durante el desarrollo del trabajo se encuentran:

- MQTT: es un protocolo que permite la comunicación entre dispositivos desarrollado por IBM. Este protocolo es considerado muy seguro debido a que utiliza un intermediario llamado MQTT Broker para el intercambio de información entre el servidor y el cliente. De esta forma, ninguna de las partes tiene acceso directo a los datos del otro, visto que solo cuentan con la información para acceder al MQTT Broker. En cambio, el cliente solicita una suscripción a un tema que le interesa (en este caso la transmisión de vídeo), y el agente se encarga de distribuir los mensajes enviados por el servidor. Está pensado especialmente para dispositivos IoT con recursos limitados y conexiones con poco ancho de banda.
- RTMP: es un protocolo diseñado específicamente para la transmisión de audio y vídeo con gran rendimiento. Esto lo logra dividiendo la información en pequeños fragmentos, cuyo tamaño puede ser negociado de manera dinámica entre el cliente y el servidor de forma que la transmisión de vídeo se adapta mejor a las condiciones de la red de la que depende la transmisión. Sin embargo, dado que se trata de un protocolo diseñado para la transmisión de vídeo y audio con el menor *delay* posible, este protocolo puede no ser el más adecuado cuando las condiciones de la red son limitadas.

2.4. Software y hardware utilizados

Para su funcionamiento, este trabajo requiere la presencia de algunos paquetes de software instalados en la máquina, los cuales son llamados directamente por el usuario cuando requiera la funcionalidad (por ejemplo, OBS Studio), o bien, que ya vienen incorporados en el código y por lo tanto, son un requisito indispensable para el trabajo y deberán encontrarse instalados. Dicho esto, es posible clasificar el software utilizado en tres categorías:

1. Paquetes de Python: son paquetes importados en el código a través del comando *import*. Cabe mencionar que todos los paquetes son de código abierto. Ejemplos de estos paquetes son:

- Numpy
- Matplotlib
- OpenCV

2. Código en Python: en especial para la implementación de la metodología YOLOV3. En internet está disponible el código fuente en formato Python (py).

3. Software completo: en particular para la retransmisión del vídeo procesado, se utilizó OBS Studio. Esta decisión responde a que, dentro del alcance del trabajo, no se definió ninguna actividad que no se encuentre directamente relacionada con la implementación de la detección de intrusos. Sin embargo, dado que se trata de un trabajo en el que se debe devolver la imagen procesada, es importante asegurarse que la retransmisión de los datos funcione de la manera correcta.

Más específicamente, las diferentes bibliotecas y paquetes de software utilizados para el desarrollo de este trabajo son:

TABLA 2.2. Software requerido para el desarrollo y la ejecución del módulo.

Paquete	Función
PIP	Gestor de bibliotecas para Python
Numpy	Cálculos matriciales
OpenCV	Manipulación de imágenes y visión por computadora
Matplotlib	Creación de visualizaciones con Python
Tensorflow	<i>Framework</i> de machine learning
Pycharm	Ambiente de desarrollo
OBS Studio	<i>Streaming</i> de vídeo
Python	Lenguaje en el que fue programado el módulo
Ubuntu	Sistema operativo

Es importante tener en cuenta que la totalidad de programas y bibliotecas utilizadas, tanto para el desarrollo, como para la ejecución del trabajo son software libre, por lo que no se requerirá adquirir ningún tipo de licencia para ponerlo en funcionamiento. En el apéndice A se encuentra la tabla completa que incluye tipo de biblioteca, licencia y versión utilizada.

De igual manera, cabe resaltar que OBS Studio es el programa de retransmisión elegido teniendo en cuenta que esta es una característica que se encuentra fuera

del alcance del trabajo y por lo tanto, cuya implementación es necesaria a través de este método mientras que hacerlo a través de software integrado en el código fuente (como FFMpeg) será parte del alcance de trabajos posteriores.

El paquete de Python ya viene instalado por defecto en la instalación de Ubuntu, por lo que no se requerirá su instalación. El resto de los paquetes, sin embargo, deberán ser instalados de acuerdo con el manual de instalación que se entrega al cliente, en el que se detalla de manera minuciosa cuál es el procedimiento que se debe seguir a fin de instalar correctamente todos los paquetes.

De igual manera, se debe tener en cuenta que dadas las restricciones de la máquina en la que se desarrolló el trabajo, la totalidad del desarrollo fue llevado a cabo en máquinas virtuales a través del uso de programas de virtualización como VMware Workstation [14]. Dentro de los requerimientos mínimos de hardware del trabajo, se tienen los siguientes:

- Procesador: AMD Ryzen 5
- Tarjeta gráfica: no requerida
- Memoria RAM: 6 GB.

Es importante tener en cuenta que los listados anteriormente son los requisitos mínimos del sistema, que van a garantizar que el trabajo se ejecute cumpliendo con los requerimientos del cliente. Sin embargo, se recomiendan las siguientes características del sistema para un rendimiento óptimo a la hora de ejecutar el módulo.

- Procesador: Intel Core i7
- Tarjeta gráfica: NVidia Tesla T4 o NVidia Titan X.
- Memoria RAM: 8 GB.

El modelo YoloV3, sobre el que está basado el desarrollo del trabajo utiliza de manera extensiva la API Cuda de NVIDIA. Por este motivo, tratar de ejecutar el módulo en tarjetas gráficas de AMD resultará en el modelo siendo llevado al procesador por Tensorflow. En consecuencia, se afirmará que el trabajo resulta incompatible con estas tarjetas gráficas. Esto fue comprobado con el uso de una tarjeta AMD Radeon Vega 8.

Capítulo 3

Diseño e implementación

En este capítulo se detallan los aspectos técnicos del desarrollo del trabajo. Incluye las consideraciones tomadas en cuenta durante el desarrollo, el detalle de las modificaciones realizadas al modelo de partida, así como el detalle del papel que juega el módulo en el sistema. De igual manera, se hace una descripción cronológica del desarrollo del módulo.

3.1. Consideraciones generales

El módulo de inteligencia artificial propuesto es una implementación del modelo YOLOv3, especialmente modificada para cumplir con las necesidades específicas del cliente, así como una serie de requisitos funcionales. Es por esto que fue necesario hacer modificaciones a los archivos originales de la biblioteca.

(Esta sección aún no se ha terminado)

3.2. Arquitectura del proyecto

3.3. Esquema del módulo

3.4. Esquema de modificaciones

3.5. Ajustes para cumplir con el rendimiento

3.6. Entrega de resultados al resto del software

Capítulo 4

Ensayos y resultados

A pesar de que esto se estipuló en la concepción inicial del proyecto, conforme se dio su avance no se hizo posible llevar a cabo un proceso formal de pruebas al módulo. Sin embargo, el cliente estuvo involucrado directamente en el proceso de verificación de la funcionalidad del módulo a través de reuniones quincenales.

4.1. Descripción del proceso de pruebas

A pesar de que esto se estipuló en la concepción inicial del proyecto, conforme se dio su avance no se hizo posible llevar a cabo un proceso formal de pruebas al módulo. Sin embargo, el cliente estuvo involucrado directamente en el proceso de verificación de la funcionalidad del módulo a través de reuniones quincenales.

4.2. Pruebas en Raspberry Pi

Desde la concepción inicial del proyecto, se estableció que se tendría dos caminos para poder desarrollarlo: Por un lado, se planteó que el módulo debía ser capaz de reconocer intrusos en varios streamings de vídeo simultáneos (camino que finalmente se vio favorecido), o bien, hacer que el módulo estuviese en capacidad de correr en equipos a bordo de los drones. Para esto, se hizo una serie de pruebas en equipos Raspberry Pi. Más concretamente, las especificaciones de los equipos sobre los que se hicieron las pruebas son:

- Modelo: Raspberry Pi
- Sistema Operativo: Raspberry Pi OS (Debian 11 – Bullseye)
- CPU quad-core ARM Cortex-A7 900MHz
- Memoria ram: 1 Gb.
- Overclocking: Desactivado

Concretamente, el modelo con el que se tuvo los problemas más importantes, especialmente a la hora de la instalación fue la versión completa de Tensorflow, requerida para la ejecución del módulo. Cabe mencionar que las placas Raspberry Pi son compatibles con la versión Lite de Tensorflow. Dicho esto, se intentaron 5 métodos para la instalación de esta biblioteca. Ninguno de ellos fue exitoso. El proceso llevado a cabo en cada una de las cinco instalaciones fue el siguiente:

- Máquina 1: Aquí irá una descripción de la instalación de la máquina
- Máquina 2

- Máquina 3
- Máquina 4
- Máquina 5

Es importante destacar que, dado que el módulo se desarrolló tomando como base el modelo YOLOv3 original (no el modelo Lite), no se cumplía con las especificaciones requeridas por este tipo de equipos, y una segunda implementación del modelo utilizando la versión YOLOv3 Lite hubiese implicado en sí, el trabajo equivalente a un proyecto completo adicional.

De igual manera atrae particular atención mencionar que la placa con la que se contaba era un modelo antiguo de Raspberry Pi (de febrero de 2015), que no cuenta con una compatibilidad tan completa como la que se da con una Raspberry Pi 4. Es, entonces, de vital importancia que, para iteraciones posteriores del desarrollo del trabajo, y en caso de que se decida continuar con el desarrollo de módulos a bordo de los drones, considerar placas más avanzadas, como Raspberry Pi 4.

4.3. Caso de Uso

Capítulo 5

Conclusiones

5.1. Resultados Obtenidos

La idea de esta sección es resaltar cuáles son los principales aportes del trabajo realizado y cómo se podría continuar. Debe ser especialmente breve y concisa. Es buena idea usar un listado para enumerar los logros obtenidos.

Algunas preguntas que pueden servir para completar este capítulo:

- ¿Cuál es el grado de cumplimiento de los requerimientos?
- ¿Cuán fielmente se pudo seguir la planificación original (cronograma incluido)?
- ¿Se manifestó algunos de los riesgos identificados en la planificación? ¿Fue efectivo el plan de mitigación? ¿Se debió aplicar alguna otra acción no contemplada previamente?
- Si se debieron hacer modificaciones a lo planificado ¿Cuáles fueron las causas y los efectos?
- ¿Qué técnicas resultaron útiles para el desarrollo del proyecto y cuáles no tanto?

5.2. Tiempos de ejecución

Acá se indica cómo se podría continuar el trabajo más adelante.

5.3. Monitoreo de resultados

Apéndice A

Tablas extendidas

TABLA A.1. Software requerido para el desarrollo y la ejecución del módulo

Paquete	Tipo	Licencia	Versión
PIP	Gestor de bibliotecas	MIT	22.0.2
Numpy	Biblioteca de Python	BSD	1.21.5
OpenCV	Biblioteca de Python	Apache 2	4.5.4
Matplotlib	Biblioteca de Python	BSD	3.6.2
Tensorflow	Biblioteca de Python	Apache 2	2.10.1
Pycharm	Software completo	Apache 2	2022.3.3 Community edition
OBS Studio	Software completo	GPLv2	28.1.2 (64 Bit)
Python	Lenguaje de programación	Python Software Foundation License Version 2	3.10.6
Ubuntu	Sistema operativo	Creative Commons CC-BY-SA version 3.0 UK	22.04.1 LTS (64 Bit)

Bibliografía

- [1] Drone AI Services. *Drones de seguridad privada ¿Cómo funcionan?* Último acceso el 28 de marzo de 2023. 2020. URL: <https://www.droneservices.com.ar/industria-4-0/drones-de-seguridad-privada/>.
- [2] Jordan Henrio y Tomoharu Nakashima. «Anomaly Detection in Videos Recorded by Drones in a Surveillance Context». En: oct. de 2018, págs. 2503-2508. DOI: [10.1109/SMC.2018.00429](https://doi.org/10.1109/SMC.2018.00429).
- [3] Joseph Redmon y col. «You Only Look Once: Unified, Real-Time Object Detection». En: *CoRR* abs/1506.02640 (2015). arXiv: [1506.02640](https://arxiv.org/abs/1506.02640). URL: <http://arxiv.org/abs/1506.02640>.
- [4] Joseph Redmon y Ali Farhadi. «YOLOv3: An Incremental Improvement». En: *CoRR* abs/1804.02767 (2018). arXiv: [1804.02767](https://arxiv.org/abs/1804.02767). URL: <http://arxiv.org/abs/1804.02767>.
- [5] Heng Zhao y col. «Mixed YOLOv3-LITE: A Lightweight Real-Time Object Detection Method». En: *Sensors (Basel)* 20.7 (2020), pág. 1861. ISSN: 1424-8220. DOI: [10.3390/s20071861](https://doi.org/10.3390/s20071861). URL: <https://doi.org/10.3390/s20071861>.
- [6] Common Objects in Context. COCO. Último acceso el 28 de marzo de 2023. URL: <https://cocodataset.org/#home>.
- [7] Tae-Hoon Yong. *Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network*. Último acceso el 11 de abril de 2023. 2020. URL: <https://www.researchgate.net/profile/Tae-Hoon-Yong/publication/342402762/figure/fig2/AS:957322327257088@1605254837269/A-CNN-architecture-modified-from-YOLOv3-with-the-modified-layers-in-Bold-CNN.ppm>.
- [8] Sairaj Neelam. *YOLO for Object Detection, Architecture Explained!* Último acceso el 11 de abril de 2023. 2021. URL: <https://medium.com/analytics-vidhya/understanding-yolo-and-implementing-yolov3-for-object-detection-5f1f748cc63a>.
- [9] Sambasivarao. K. *Non-maximum Suppression (NMS)*. Último acceso el 11 de abril de 2023. 2019. URL: <https://towardsdatascience.com/non-maximum-suppression-nms-93ce178e177c>.
- [10] Uri Almog. *YOLO V3 Explained*. Último acceso el 11 de abril de 2023. 2020. URL: <https://towardsdatascience.com/yolo-v3-explained-ff5b850390f>.
- [11] Weidong Kuang. *Fundamentals of deep learning: a step-by-step guide Chapter 12 Object Detection*. Último acceso el 11 de abril de 2023. 2023. URL: <https://faculty.utrgv.edu/weidong.kuang/book/chapter12%20object%20detection.pdf>.
- [12] Manish Chablani. *YOLO — You only look once, real time object detection explained*. Último acceso el 11 de abril de 2023. 2017. URL: <https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006>.

- [13] Jonathan Hui. *Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3*. Último acceso el 11 de abril de 2023. 2018. URL: <https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-28b1b93e2088>.
- [14] VMware. *VMware Workstation Player*. Último acceso el 11 de abril de 2023. 2023. URL: <https://www.vmware.com/products/workstation-player.html>.