



Universidad
Zaragoza

Trabajo Fin de Grado

Influencia de la perspectiva en el
reconocimiento de objetos con redes
neuronales

Influence of perspective in object recognition
with neural networks

Autor/es

César José, Pardos Rivera

Director/es

David, Morilla Cabello

ESCUELA DE INGENIERÍA Y ARQUITECTURA
2023

RESUMEN

Las redes neuronales han demostrado ser herramientas versátiles con aplicaciones en una amplia variedad de campos. Una de estas aplicaciones es el reconocimiento de objetos a través de la segmentación semántica. La posición de la cámara se convierte en un factor importante a considerar ya que puede impactar significativamente en la capacidad con la que se reconoce y procesa el objeto. Sin embargo, dado que estas redes se entrenan a partir de un conjunto limitado de datos, hay diversos factores a tener en cuenta. En aplicaciones reales, donde la cámara puede situarse en el espacio, su posición puede determinar la capacidad con la que se reconocen los objetos.

En este trabajo se va a realizar un estudio de la influencia que tienen la posición y la orientación de la cámara con respecto a los objetos y el resto del entorno para evaluar la precisión de la segmentación semántica lograda por las redes neuronales.

Mediante el uso de un simulador fotorrealista, se generan diversas configuraciones de posición y orientación de la cámara en un entorno virtual donde se capturan imágenes y procesan mediante redes neuronales entrenadas para realizar tareas de segmentación semántica. Finalmente se realizan los cálculos de la métrica *Intersection over Union* comparando las áreas de superposición entre las segmentaciones predichas y las segmentaciones reales.

Los resultados obtenidos a partir de la métrica de evaluación permiten analizar cómo la posición de la cámara afecta la precisión y la coherencia de las predicciones de las redes neuronales. Estos resultados proporcionan información relevante sobre qué posiciones de la cámara pueden dificultar o mejorar la interpretación de la información visual por parte de las redes.

PALABRAS CLAVE: Redes neuronales, posición de cámara, *Intersection over Union*, Segmentación semántica

ABSTRACT

Neural networks have proven to be versatile tools with applications in a wide variety of fields. One of these applications is object recognition through semantic segmentation. The position of the camera becomes an important factor to consider as it can significantly impact the ability with which the object is recognised and processed. However, since these networks are trained on a limited set of data, there are several factors to take into account. In real applications, where the camera can be placed in space, its position can determine the ability with which objects are recognised.

In this paper we will study the influence of the position and orientation of the camera with respect to objects and the rest of the environment to evaluate the accuracy of semantic segmentation achieved by neural networks.

Using a photorealistic simulator, various camera position and orientation configurations are generated in a virtual environment where images are captured and processed by neural networks trained to perform semantic segmentation tasks. Finally, calculations of the Intersection over Union metric are performed by comparing the areas of overlap between the predicted segmentations and the real segmentations.

The results obtained from the evaluation metric allow us to analyse how the camera position affects the accuracy and consistency of the neural networks' predictions. These results provide relevant information on which camera positions can hinder or improve the interpretation of visual information by the networks.

AGRADECIMIENTOS

Este trabajo y final de etapa se lo dedico a mis padres por estar siempre ahí.

RESUMEN	2
ABSTRACT	3
AGRADECIMIENTOS	4
ÍNDICE DE FIGURAS	7
1.Introducción	9
1.1 Motivación	9
1.2 Objetivo y alcance	9
1.3 Estructura de la memoria	12
2. Herramientas	13
2.1 Python	13
2.1.1 Pycharm	14
2.2 Unreal y Airsim	14
2.3 Red neuronal entrenada: Deeplab V3 Resnet 50	14
3.Fundamentos	15
3.1 Redes neuronales	15
3.1.1 Definición y funcionamiento	15
3.1.2 Capacidad de aprendizaje	17
3.1.3 Redes neuronales convolucionales	17
3.1.4 Segmentacion semántica	18
3.2 Intersection Over Union	19
4. Metodología	20
4.1 Aprendizaje previo	20
4.2 Preparación del entorno de trabajo	21
4.3 Obtención de imágenes	21
4.4 Procesamiento de imágenes y evaluación	23
4.5 Diseño de experimentos	24
5. Experimentos	25
5.1 Estudio sobre la influencia del punto de vista	27
5.2. Estudio de oclusiones parciales en el punto de vista	29
5.3 Estudio de complejidad en diferentes escenarios:	30
5.4 Estudio de la influencia de la posición no convencional del objeto de estudio.	31
5.5 Estudio de complejidad de la escena	32

6.RESULTADOS DE LOS EXPERIMENTOS	35
6.1 Estudio sobre la influencia del punto de vista	35
6.2 Estudio de oclusiones parciales en el punto de vista	40
6.3 Estudio de complejidad en diferentes escenarios	43
6.4 Estudio de la influencia de la posición no convencional del objeto de estudio.	45
6.5 Estudio de complejidad de la escena	46
7. CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURAS	48
REFERENCIAS	50

Índice de figuras

Figura 1: Robot recolectando tomate guiado con cámara de visión.	8
Figura 2: Robot SPOT realizando una inspección arqueológica	9
Figura 3 : Cronograma de las diferentes tareas del desarrollo de este trabajo.	11
Figura 4: Esquema de una red neuronal artificial .	15
Figura 5: Esquema funcionamiento de una red neuronal	15
Figura 6: Esquema de una red neuronal convolucional	17
Figura 7: Ejemplo de segmentación semántica	17
Figura 8 : Métrica de Intersección Over Union	18
Figura 9 : Flujo de trabajo utilizado para la obtención de imágenes	20
Figura 10: Imagen RGB a la izquierda e imagen GT a la derecha	22
Figura 11 : Flujo de trabajo utilizado para el procesamiento de imágenes	23
Figura 12: Comparación imagen Ground truth y red neuronal con el objeto de estudio perro.	24
Figura 13: Posiciones de la cámara en el movimiento esférico	26
Figura 14: Escenario urbano a la izquierda y escenario rural a la derecha.	27
Figura 15: Diferentes ángulos de inclinación para un ángulo de orientación de 180º	28
Figura 16: Comparación de las diferentes distancias en imagen RGB a 1,2,3 y 7 metros.	28
Figura 17: Diferentes ángulos de posición para un ángulo de inclinación de 0º	29
Figura 18: Comparación de las diferentes grado de oclusión en RGB: A la izquierda alta oclusión, a la derecha sin ocluir, en el centro parcialmente ocluido.	30
Figura 19: Escenario urbano y rural de la silla a 3 metros.	30
Figura 20: Imagen de los 3 objetos en posición no convencional.	31
Figura 21: Imagen del objeto de estudio caballo y en el fondo el objeto barco.	32
Figura 22: Relación entre IoU y posición para 0º ángulo de inclinación constante.	33
Figura 23: Imágenes RGB, GT y semántica para valores de ángulo de inclinación 0º, 30º , 60º y 90º.	34
Figura 24: Imagen sobre la influencia de la cercanía y lejanía.	36
Figura 25: Posiciones de estudio para resultados de IoU altos. Para el perro 180º de punto de vista, 0º para la bicicleta y 180º para la silla.	37

Figura 26: Posiciones de estudio para resultados de IoU altos. Para el perro 180º de punto de vista, 0º para la bicicleta y 180º para la silla.	37
Figura 27: Relación entre IoU y posición para 0º ángulo de inclinación constante.	38
Figura 28 : Grado de oclusión para 0º ángulo de inclinación según el objeto de estudio.	39
Figura 29 : Comparación de los diferentes grados de oclusión en RGB, GT y procesada	40
Figura 30 : Relación entre la distancia e IoU para Silla	40
Figura 31 : Relación entre la distancia e IoU para Bicicleta	41
Figura 32 : Relación entre la posición convencional y no convencional	42
Figura 33 : Posición no convencional para objeto de estudio caballo.	43
Figura 34 : Posición no convencional para objeto de estudio bicicleta.	43
Figura 35 : Relación entre la distancia e IoU para cada objeto	44
Figura 36: Imagen RGB, imagen GT y imagen semántica procesada por red neuronal.	45

1. Introducción

1.1 Motivación

Las redes neuronales, inspiradas en el cerebro humano, han revolucionado la manera en que las máquinas comprenden y analizan el mundo visual que las rodea. Al emplear capas interconectadas para procesar datos visuales, las redes neuronales han logrado avances en el reconocimiento de objetos, detección de rostros, segmentación semántica y más. Ejemplo de ello es en el campo de la automoción, las redes neuronales convolucionales han demostrado su eficacia en la conducción autónoma, reconociendo y clasificando todo tipo de señales y objetos [14]. Otro campo de aplicación es en la agricultura con la recolección de fruta y verdura, como se observa en la Figura 1 [18] o bien, en arqueología con la creación de mapas digitales mediante la visión por computación [19], se observa en la Figura 2 .



Figura 1: Robot recolectando tomate guiado con cámara de visión [18].

La motivación detrás de este proyecto ha radicado en analizar el complejo papel que juega la posición y orientación de la cámara en la precisión de la identificación de objetos. El uso de redes neuronales entrenadas ha ofrecido una capacidad para analizar patrones que pueden escapar a la percepción humana. Explorar cómo las redes neuronales capturan y procesan información sobre la posición podría tener implicaciones significativas en áreas como la visión artificial, la robótica y la seguridad.



Figura 2: Robot SPOT realizando una inspección arqueológica [19].

Al abordar esta investigación, no solo se ha ampliado el conocimiento en el campo de la inteligencia artificial y la visión por computador, sino que también se estará contribuyendo a resolver desafíos prácticos en el mundo real. Comprender cómo una red neuronal entrenada ha mejorado el reconocimiento de objetos en diversas posiciones de cámara podría tener aplicaciones en vehículos autónomos, predicción del rendimiento de cultivos y en el control de robots.

A la influencia de la posición se suman otros factores externos para su análisis, tales como el escenario del objeto. Durante el procesamiento de la imagen, los elementos que rodean al objeto de estudio pueden llegar a ocasionar desde falsos positivos hasta la no detección de ciertas zonas, por lo que podría llevar a errores en la identificación de la silueta del objeto. Otro factor de gran interés que se explora en este proyecto es la oclusión parcial del objeto de estudio, que puede presentar un desafío para la capacidad de detección de la red neuronal. Todos estos factores mencionados y otros muchos añaden un gran potencial a este proyecto e investigaciones futuras.

1.2 Objetivo y alcance

El principal objetivo del trabajo es investigar la influencia que tiene la posición desde donde se toma una imagen, así como factores externos como el contexto o las oclusiones en las cámaras de visión artificial mediante redes neuronales entrenadas. El reconocimiento de un objeto varía significativamente entre una vista frontal y una vista desde un perfil lateral, además de que la inclinación de la captura y la perspectiva

también desempeñan un papel crucial en este proceso. Otros factores importantes son el entorno que rodea al objeto en cuestión, la posible oclusión que pueda afectar en la visibilidad del objeto de estudio, la distancia desde donde realiza la imagen o la posición espacial del objeto.

Este trabajo ha supuesto para el autor una toma de contacto y aprendizaje del mundo de la programación informática y redes neuronales, que ha resultado muy interesante para su formación académica. Además, explorar cómo la posición de la cámara afecta el rendimiento de la red neuronal ha enriquecido su comprensión en visión por computadora y potencialmente abre puertas hacia futuras investigaciones interdisciplinarias.

Con objeto de lograr un proyecto consistente este objetivo principal se ha descompuesto en los siguientes hitos:

- Introducción y aprendizaje a la programación en el lenguaje Python y diferentes librerías para el deep learning como NumPy y Pytorch.
- Introducción y adaptación al trabajo con redes neuronales y segmentación semántica.
- Diseño y creación de los diferentes escenarios fotorrealistas de trabajo desarrollados en Unreal.

Además en la parte experimental se han logrado los siguientes objetivos para su desempeño:

- Estudio y diseño sobre ejecución y movimiento de la toma de imágenes experimentales.
- Desarrollo del código empleado en Python para la captura de imágenes.
- Cálculo de la métrica de Intersection Over Union (IoU) para la obtención de los resultados.

En la Figura 3, se observa un cronograma del reparto de las diferentes tareas para llegar a cabo a los objetivos nombrado

NOMBRE DE LA TAREA	Nº DE HORAS	dic-22	ene-23	feb-23	mar-23	abr-23	may-23	jun-23	jul-23	ago-23
Investigación preliminar	35									
Aprendizaje redes neuronales	15									
Instalación programas	15									
Análisis de los objetivos	5									
Formación básica	125									
Tutoriales Python	50									
Propuesta de TFG	5									
Parte experimental	70									
Desarrollo del código	30									
Estudio de los experimentos	15									
Diseño de los experimentos	15									
Extracción de las imágenes	20									
Análisis de los resultados	20									
Documentación técnica	65									
Escritura y refinado documentos	30									
Memoria	20									
Presentación	10									
Ensayo de la presentación	5									
Tiempo total	300									

Figura 3 : Cronograma de las diferentes tareas del desarrollo de este trabajo.

1.3 Estructura de la memoria

En este apartado, se proporciona una introducción concisa donde se expone la motivación que subyace en este proyecto, además de presentar los objetivos principales que se abordan en él. Junto a esto, se perfila la estructura global de la memoria para otorgar una visión temporal del desarrollo de lo contenidos en la Figura 3.

Los siguientes dos apartados, el 2 y el 3, toman un enfoque más profundo donde se introducen las herramientas empleadas y los fundamentos empleados en el proceso. En el segundo apartado, se destaca la utilidad de herramientas como PyCharm y Airsim, y se establece una introducción a la red neuronal entrenada que desempeñará un rol central en los experimentos. En el tercer apartado, se explican los fundamentos esenciales para el entendimiento de este proyecto. Se aborda la estructura de una red neuronal, su proceso de aprendizaje y se explica con detalle cómo la red neuronal opera bajo el principio de segmentación semántica. Además, se acuña la métrica utilizada para evaluar su capacidad de reconocimiento.

A continuación, traspasando los límites de la metodología, se presenta de manera clara y concisa el método desarrollado y aplicado en los experimentos.

Los apartados 5 y 6 son piedras angulares de la parte experimental del proyecto. En el primero, se detalla el propósito específico de cada experimento y se les dota de descripciones minuciosas para entender sus contextos. En el segundo, se efectúa un análisis exhaustivo de los resultados obtenidos en estos experimentos, cerrando así el ciclo experimental.

Para concluir, en el apartado 7, se extraen conclusiones fundamentales que emergen de los resultados de los apartados 5 y 6. Asimismo, se abre la puerta a futuras líneas de investigación que se desprenden de este trabajo.

2. Herramientas

2.1 Python

Python es el lenguaje de programación elegido para desarrollar el código de los experimentos. Se destaca por su sintaxis legible y su enfoque en la legibilidad del código. Ampliamente utilizado en desarrollo web, análisis de datos y automatización.

Python cuenta con una amplia variedad de librerías que abarcan una amplia gama de aplicaciones. Algunas de las principales librerías con las que se ha trabajado son:

NumPy: Es una librería fundamental para la computación numérica en Python. Proporciona soporte para vectores multidimensionales y funciones matemáticas de alto rendimiento, lo que es esencial para la manipulación de datos y la implementación eficiente de algoritmos numéricos [1].

PyTorch: Estas son dos de las librerías más populares para el aprendizaje profundo (deep learning). En ellas se permite construir, entrenar y desplegar redes neuronales y modelos de aprendizaje profundo en una variedad de aplicaciones. Ambas se caracterizan por trabajar con tensores que son estructuras de datos multidimensionales que fluyen (flow) a través de operaciones matemáticas [1].

Pillow (PIL Image) es una biblioteca de procesamiento de imágenes en Python. Se ha diseñado para trabajar con imágenes en varios formatos y realizar una amplia gama de operaciones de manipulación y procesamiento de imágenes, así como abrir, editar, crear y guardar imágenes. En el código realizado, se utiliza para trabajar con las imágenes de los experimentos [1].

2.1.1 Pycharm

PyCharm es un entorno de desarrollo integrado y se utiliza para programar en Python el código del proyecto. Se ofrece un editor avanzado, herramientas de depuración visuales, gestión de proyectos, integración con sistemas de control de versiones y soporte para marcos de trabajo y librerías de Python. [2] En el código que se ha realizado destacan las siguientes funciones de valor como: transformación de la imagen RGB a imagen de tensor de números, bucle para la obtención de imágenes, configuración de clases del simulador entre otras.

2.2 Unreal y Airsim

Unreal Engine es una herramienta desarrollada por Epic Games ampliamente utilizada en la industria de videojuegos. Ofrece una plataforma completa para diseñar y desarrollar experiencias interactivas, enfocándose en crear mundos digitales detallados para juegos y aplicaciones en tiempo real. Sus características incluyen un motor gráfico de alto rendimiento, simulación física realista y un avanzado sistema de iluminación. Además, la versatilidad de Unreal Engine hace que sea la base de AirSim, una extensión desarrollada por Microsoft [3].

AirSim se especializa en la simulación de vehículos autónomos y drones, utilizando la capacidad de simulación y los gráficos realistas de Unreal Engine. Esto ha permitido a investigadores y desarrolladores probar algoritmos y sistemas en un entorno virtual antes de experimentar en el mundo real [4].

2.3 Red neuronal entrenada: Deeplab V3 Resnet 50

Como el propósito central del proyecto radica en el análisis de la influencia de la posición de la cámara, se ha optado por dirigir la atención a la parte experimental. En este sentido, se ha tomado la decisión de emplear una red neuronal que ya cuenta con un proceso de entrenamiento previo. Además este tipo de redes se han entrenado por grandes compañías con capacidad de cómputo, por lo que se han centrado los esfuerzos en el estudio de los experimentos.

La elección recae en la red neuronal denominada Deeplab V3 Resnet 50, la cual opera a través de una arquitectura convolucional. Si nos adentramos un poco en la

terminología, el número 50 asociado a Resnet 50 se refiere a la cantidad de capas de neuronas que componen su estructura de codificación (encoder).

Esta red está configurada para trabajar mediante segmentación semántica, es decir, clasifica los píxeles de una imagen según las categorías que representan. A diferencia de otras redes neuronales, como YOLO, en las que su forma de reconocimiento se realiza mediante la detección del contorno que rodea al objeto.

La red ha sido entrenada utilizando el conjunto de datos COCO (Common Objects in Context), que provee una amplia variedad de imágenes para los algoritmos de detección. Dada la extensión de COCO, se ha optado por la configuración PASCAL LABEL, la cual simplifica la tarea de detección al centrarse en 21 clases predefinidas, como "barco", "flores" y "silla".

3. Fundamentos

3.1 Redes neuronales

3.1.1 Definición y funcionamiento

Una red neuronal se define como un sistema capaz de establecer una relación entre la entrada y la salida mediante el uso de coeficientes y variables interrelacionadas. Al interconectar estas redes neuronales, se logra emular un sistema nervioso y se diferencia de la computación tradicional al no especificar los algoritmos, aprendiéndolos de forma automática en base a los datos.

Las redes neuronales artificiales imitan el funcionamiento del cerebro humano al procesar la información de manera paralela y poseer la capacidad de aprender, pudiendo así generalizar situaciones que no se incluyeron en el proceso de entrenamiento [6].

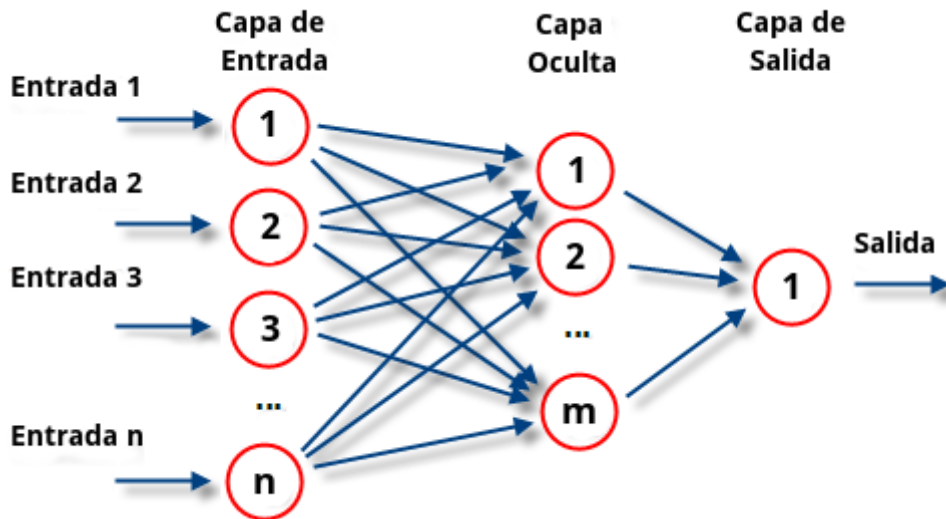


Figura 4: Esquema de una red neuronal artificial [15].

En el esquema de una red neuronal ilustrado en la Figura 4, se pueden apreciar las n entradas del sistema, cada una de las cuales se encuentra conectada a una capa oculta y posteriormente se unen a la capa de salida, ambas con diversos valores de ponderación establecidos. Los valores paramétricos de estas ponderaciones entre capas han sido previamente optimizados o “aprendidos” en la red neuronal durante una fase de entrenamiento.

A continuación en la Figura 5 tenemos una red neuronal en particular, donde forma el vector de entrada mediante el sumatorio de las variables $[x_0, x_1, x_2]$ combinado con los coeficientes de entrada $[w_0, w_1, w_2]$. El resultado del sumatorio se introduce a una función de activación para posteriormente obtener el vector de salida resultante.

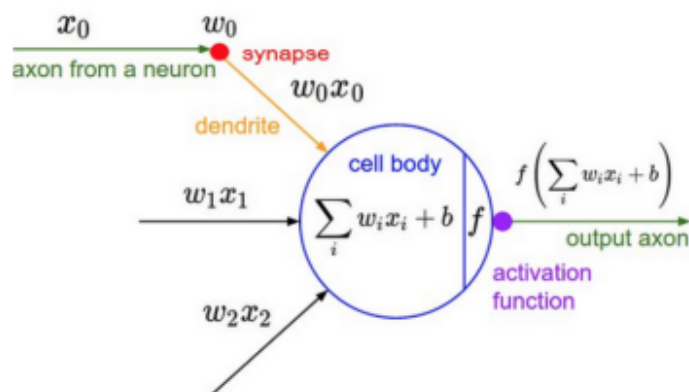


Figura 5: Esquema funcionamiento de una red neuronal [16].

3.1.2 Capacidad de aprendizaje

La capacidad de aprendizaje que se adapta según su experiencia, que constituye una de las características más interesantes de las redes neuronales. Es decir, estas redes adquieren la habilidad de realizar determinadas tareas a través de un entrenamiento con ejemplos similares a los que se desea obtener en la salida [13].

Su adaptabilidad y dinamismo viene dada por la capacidad de autoajuste de los parámetros y coeficientes que conforman el sistema de neuronas. Durante el proceso de aprendizaje, los enlaces ponderados de las neuronas se ajustan con el objetivo de alcanzar resultados específicos. Una red neuronal prescinde de la necesidad de un algoritmo para resolver un problema, ya que es capaz de generar su propia distribución de pesos en los enlaces mediante el aprendizaje [7].

3.1.3 Redes neuronales convolucionales

Las redes neuronales convolucionales son un tipo de arquitectura de redes neuronales artificiales especialmente relacionadas con la visión artificial por computación. Las aplicaciones más destacadas son en el campo de la detección de objetos, reconocimiento de rostros y el procesamiento de lenguaje natural [13]. La forma de trabajar de este tipo de redes está claramente diferenciada en 3 operaciones:

En la primera operación se produce un proceso de convolución, donde se aplican filtros a la imagen de entrada. Se puede definir ese filtro como un conjunto de matrices de números que tienen unos valores establecidos iniciales y que recorren todos los píxeles de la imagen. Según los diferentes valores del filtro, también llamado kernel, se obtiene algo parecido a la imagen inicial pero a su vez la imagen es completamente diferente a ella.

En la segunda operación se produce el agrupamiento. Obtenido el output de la primera capa, resultará ser el nuevo input de esta segunda capa, en el que se reduce la dimensión de la imagen filtrada para aumentar el número de mapas de características (patrones).

Finalmente, en la tercera operación, como ocurre en la anterior capa, el output de la segunda capa pasa a ser el input de la tercera capa, donde se produce la activación de todos los patrones devolviendo como resultado la imagen definitiva [8]. En la Figura 6 se puede observar el proceso que se ha descrito.

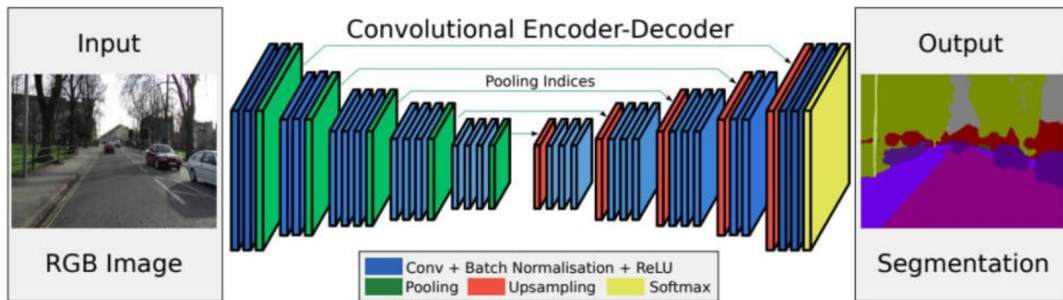


Figura 6: Esquema de una red neuronal convolucional[19].

3.1.4 Segmentación semántica

La segmentación semántica es una técnica avanzada de procesamiento de imágenes y objetos en la visión por computador, que consiste en otorgar una etiqueta o categoría a cada píxel de la imagen. La cantidad de clases a las que se pueda clasificar dependerá de cómo haya sido entrenada la red neuronal. La segmentación semántica tiene dos etapas marcadas [9]:

La primera etapa se produce la codificación de la imagen inicial, en el que se reduce las dimensiones de la imagen mediante convoluciones. En la segunda etapa, ocurre el proceso inverso a la primera etapa, se decodifica mediante el uso de procesos convolucionales, hasta obtener una imagen con las mismas dimensiones que en la primera. Uno de los aspectos más interesantes que tiene la segmentación semántica es la conexión entre codificación y decodificación, cuando esta se encuentra en el mismo nivel convolucional [13]. En la Figura 7, se puede observar el resultado de un proceso de segmentación semántica.



Figura 7: Ejemplo de segmentación semántica [5].

3.2 Intersection Over Union

Intersection over Union (IoU) es una métrica comúnmente utilizada para evaluar la precisión de modelos de detección y segmentación en tareas de visión por computadora, como detección de objetos y segmentación semántica. En esencia, Intersection over Union compara la superposición entre la región predicha por el modelo y la región real o etiquetada en un conjunto de datos. Esta métrica es especialmente valiosa cuando se trata de medir el rendimiento de algoritmos de detección y segmentación, ya que proporciona una medida cuantitativa de la calidad de las predicciones. [10]

En primer lugar, para el cálculo de IoU se obtiene el área de intersección entre la región predicha y la región etiquetada. Luego, se calcula el área de unión de ambas regiones. Finalmente, se divide el área de intersección entre el área de unión para obtener el valor del IoU, que oscila entre 0 y 1. Un IoU cercano a 1 indica una alta coincidencia entre la predicción y el objeto real, mientras que un IoU cercano a 0 indica una pobre superposición y una mala predicción. [11] [12]

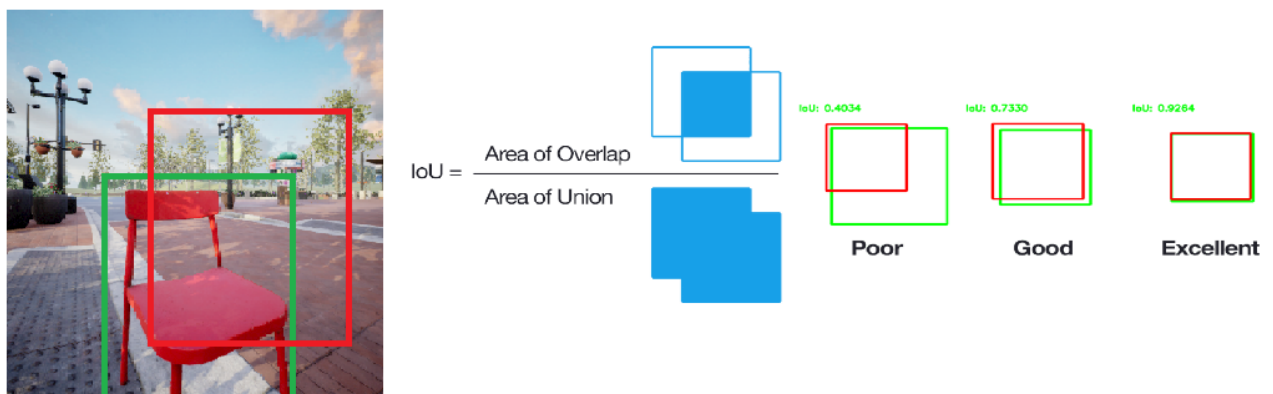


Figura 8 : Métrica de Intersección Over Union

En la Figura 8, vemos aplicado al caso particular con el objeto silla, en el que la superficie cuadrada de color verde es la que reconoce la modelo red neuronal entrenada y la superficie cuadrada de color rojo representa el caso real.

4. Metodología

En el proyecto realizado se aborda la investigación de cómo la posición y orientación de la cámara afecta en el proceso de reconocimiento de objetos. Para llevar a cabo este estudio, se ha seguido una metodología compuesta por varias etapas.

4.1 Aprendizaje previo

Dado que el autor ha carecido de conocimientos previos en el ámbito de trabajo de este proyecto, la etapa inicial ha consistido en la realización de diversos tutoriales relacionados con Python y Unreal, dos pilares fundamentales para el desarrollo de la investigación.

La principal fuente de conocimiento se ha encontrado en línea, concretamente en la página Aprende con Alf [21]. En esta plataforma, se ha encontrado un enfoque gradual y estructurado que ha permitido la adquisición de conocimientos desde sus cimientos, avanzando metódicamente a través de la estructura del sitio. Además de la teoría, se encontraron ejercicios con soluciones categorizadas por dificultad y área de aplicación en la programación en Python. En esta plataforma se aborda desde los conceptos básicos como condicionales, bucles, listas, tuplas y diccionarios hasta las áreas más avanzadas como funciones, módulos y, en particular, las librerías críticas para el proyecto, como numpy o matplotlib.

Simultáneamente, otra página [22] se puede considerar valiosa en el aprendizaje, es una plataforma similar a la anterior. Aquí, nuevamente, se encontraron abundantes ejercicios que cubrieron una amplia gama de temas en Python, permitiendo una formación integral y versátil.

Adicionalmente, ante la presencia de desafíos y problemas sin solución aparente, la herramienta Chat GPT se ha empleado como recurso de depuración. Su capacidad para proporcionar orientación y soluciones ha resultado fundamental en la resolución de los problemas más específicos.

En lo que respecta a la familiarización con Unreal, se ha recurrido a múltiples tutoriales en YouTube para maximizar el aprovechamiento de esta herramienta. La consulta de información específica en Google también ha sido una práctica recurrente para resolver dudas sobre el uso de herramientas concretas. Estas estrategias fueron esenciales para dominar las capacidades de Unreal y aplicarlas efectivamente en el proyecto.

4.2 Preparación del entorno de trabajo

Las instalaciones de PyCharm y Unreal han resultado ser procedimientos simples y muy intuitivos. Las versiones empleadas para estos programas han sido PyCharm 2022.3.3 y Unreal 4.27.0.

Sin embargo, en la instalación de Airsim se ha necesitado adicionalmente la configuración en Visual Studio del archivo setting.json. Para ello se han necesitado de una guía de instalación [23]. Adicionalmente, durante la instalación del propio programa Airsim se han presentado dificultades debido a su interacción con Windows, específicamente porque el nombre de usuario contenía un carácter especial. Para solventar esta situación, se ha tenido que crear una nueva cuenta de usuario en Windows, lo que se tuvo que reinstalar los programas de trabajo anteriores.

Adicionalmente, durante la configuración del entorno de trabajo, se ha llevado a cabo la descarga de modelos 3D de fuentes disponibles en páginas web de licencia libre. Estos modelos se incorporaron con el propósito de modificar el escenario según cada caso experimental.

4.3 Obtención de imágenes

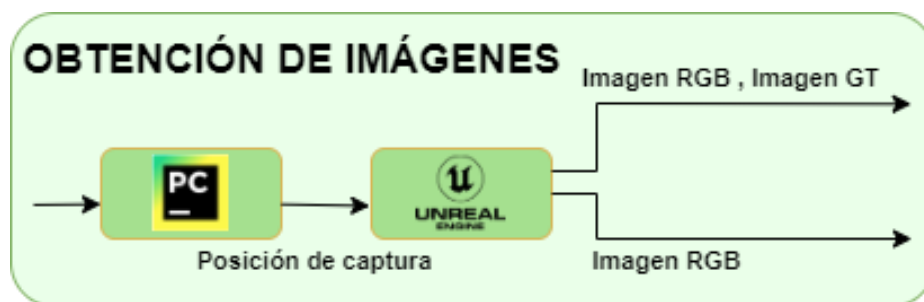


Figura 9 : Flujo de trabajo utilizado para la obtención de imágenes

En el proceso de obtención de imágenes, que se resume en la Figura 9, se han llevado a cabo diversas configuraciones en el escenario del simulador fotorrealista. Estas configuraciones abarcan desde la disposición de objetos que obstruyen la visión hasta cambios completos de escenario e incluso la inclusión de objetos que carecen de coherencia contextual con la escena.

El propósito central del código desarrollado reside en la obtención de dos tipos de imágenes: una en imagen RGB y otra en imagen semántica Ground Truth (GT). Esto se realiza desde una variedad de puntos de vista, tal como se puede observar en la Figura 10. Para alcanzar este objetivo, se empleó un enfoque automatizado que generaba posiciones de captura a lo largo de un grid, abarcando diferentes ángulos en torno al objeto de estudio.



Figura 10: Imagen RGB a la izquierda e imagen GT a la derecha

Al ejecutar el código, el proceso se inicia ajustando la gama de colores en el simulador. Este ajuste garantiza que el objeto de estudio se visualice con el color deseado. A continuación, el código entra en un ciclo, solicitando en cada iteración una imagen en formato RGB y otra en imagen semántica GT. Estas imágenes se adquieren para cada posición y ángulo de inclinación definidos en el grid. Una vez obtenidas, se almacenan en una carpeta específica para su posterior procesamiento. La cantidad de imágenes generadas depende en última instancia de la configuración del grid y del tipo de imagen que el proceso requiere.

4.4 Procesamiento de imágenes y evaluación

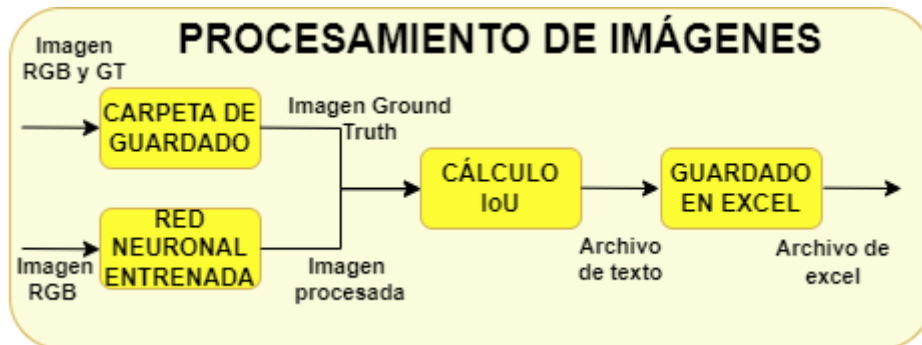


Figura 11 : Flujo de trabajo utilizado para el procesamiento de imágenes

Tras la adquisición de las imágenes en Unreal, se procede al procesamiento de las imágenes en RGB y GT, donde se observa en la Figura 11, cuyo objetivo principal es el cálculo de la métrica de evaluación IoU (Intersection over Union). En el inicio de este proceso, se lleva a cabo la carga de la red neuronal previamente entrenada en su modo de evaluación. Esta preparación es crucial para asegurar que, al abrir la imagen en imagen RGB, la red neuronal pueda procesarla de manera adecuada. Una vez completado este paso, se procede a la toma de la imagen Ground Truth (GT) y la imagen resultante del proceso neuronal, como se observa en la Figura 11, estas dos imágenes se emplean en el cálculo del IoU. El resultado de este cálculo es un valor numérico que proporciona una medida cuantitativa de la precisión del proceso de segmentación.

Debido a que el proyecto abarca un número considerable de experimentos, se ha tomado la decisión de almacenar estos resultados en una carpeta externa. Este enfoque se adopta para facilitar un análisis más completo y profundo de los resultados obtenidos. Al guardar los resultados en un archivo separado, se fomenta una revisión detallada y se posibilita la comparación entre distintos experimentos en el estudio.

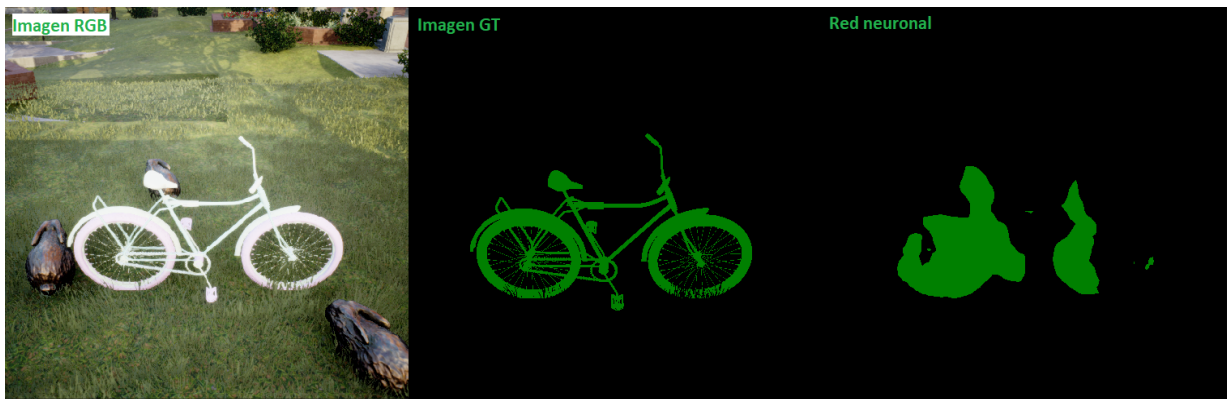


Figura 12: Comparación imagen RGB, Ground truth y red neuronal con el objeto de estudio bicicleta.

4.5 Diseño de experimentos

Antes del comienzo experimental de la captura de imágenes, se ha realizado un estudio acerca de los potenciales movimientos de cámara para realizar y los escenarios de mayor interés. Se han elegido esos movimientos en particular debido a su gran capacidad para llegar a los diversos puntos de vista que puede llegar el ojo humano. No se ha llegado a realizar un movimiento desde debajo del objeto debido a que no tiene una utilidad general porque no se suele utilizar ese punto de vista.

En la etapa experimental, en primer lugar se ha empleado el entorno de desarrollo PyCharm para la ejecución del código desarrollado específicamente para este proyecto. Este código se ha diseñado para conectarse al simulador fotorrealista y controlar el movimiento de la cámara, permitiendo capturar imágenes desde diferentes ángulos y perspectivas. Cada una de estas imágenes se convierte en su correspondiente Ground Truth (verdad de referencia) con el fin de tener una base con la cual comparar los resultados del sistema propuesto. Una vez obtenidas las imágenes, se procede a procesar una de ellas utilizando una red neuronal previamente entrenada para la tarea de reconocimiento de objetos. Esta red neuronal es capaz de identificar objetos en la imagen y proporcionar resultados de reconocimiento. La otra imagen, la cual se considera la Ground Truth de esta prueba, actúa como el estándar de comparación.

Para evaluar la similitud entre los resultados de la red neuronal y la Ground Truth original, se emplea una métrica específica. Esta métrica cuantifica la similitud y permite establecer qué tan precisas son las predicciones de la red neuronal en comparación con la verdad de referencia.

Los valores de similitud calculados junto a la posición y ángulo que corresponde se registran y almacenan en un archivo de texto (formato .txt). Estos valores, que reflejan

la relación entre las imágenes procesadas y las Ground Truth, se procesan posteriormente para generar un archivo en formato Excel. Esta hoja de cálculo permite una presentación más organizada y accesible de los resultados, lo que facilita su análisis y posterior obtención de gráficas.

5. Experimentos

En este apartado describe los experimentos concretos que se han realizado en el TFG siguiendo la metodología planteada en el capítulo anterior. Los experimentos se han llevado a cabo en un ordenador equipado con un procesador i7-9750H y una GTX2070 y la red neuronal empleada en estos experimentos ha sido Deeplab V3 Resnet 50.

En todos los experimentos se ha seguido el mismo procedimiento para la toma de imágenes, difiriendo únicamente en la distancia a la que se han realizado las capturas y el objeto de estudio. Como se muestra en la Figura 13, el movimiento de la cámara recorre las 6 posiciones de las flechas azules y, en cada una de ellas, se toman 4 imágenes descritas por las flechas amarillas. Este movimiento se realiza para tener una perspectiva desde diferentes ángulos de inclinación, haciendo un total de 24 imágenes.

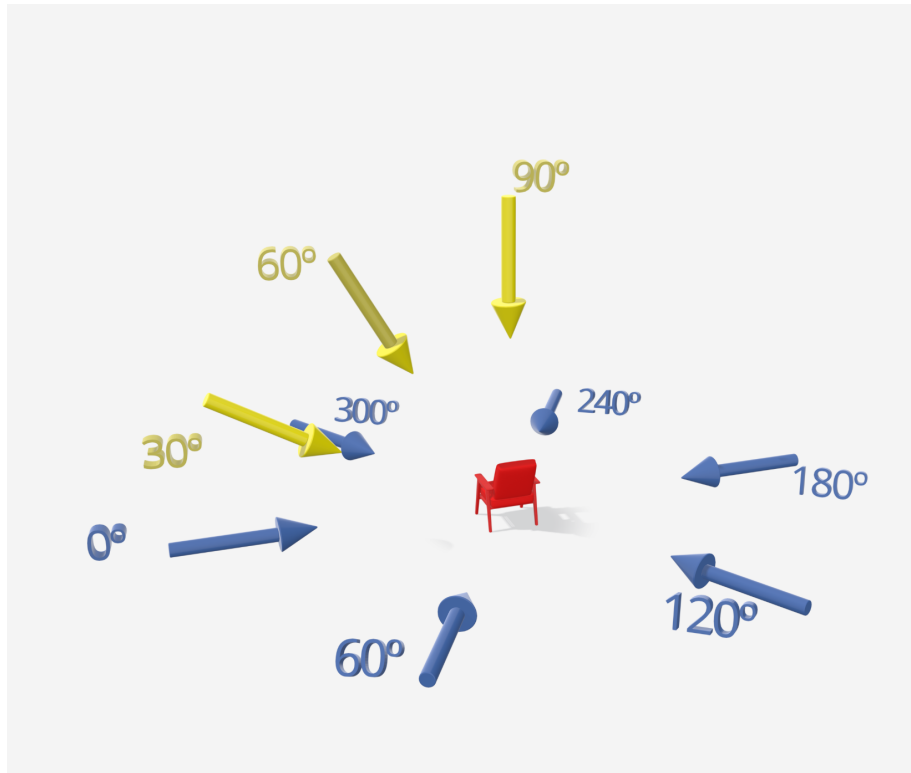


Figura 13: Posiciones de la cámara en el movimiento esférico

En cuanto al escenario, gracias a la posibilidad que da la utilidad de un simulador fotorrealista se plantearon diferentes lugares pero se ha optado por dos en particular, uno de carácter urbano y otro de carácter rural. Su elección se debe a la frecuencia y similitud que tienen con los escenarios de la vida real. Ambos se encuentran en el entorno virtual que podemos encontrar en la biblioteca de Unreal de manera gratuita bajo el nombre 'Demo_Environment'. En la Figura 14 se muestra un ejemplo de imagen capturada en cada uno de los dos escenarios.



Figura 14: Escenario urbano a la izquierda y escenario rural a la derecha.

5.1 Estudio sobre la influencia del punto de vista

En este experimento, se han adquirido imágenes desde cuatro distancias distintas, como se puede ver en la Figura 16 , que abarcan desde proximidades hasta posiciones más alejadas del objeto en estudio. La elección de realizar este experimento surge de la hipótesis de que tanto las distancias cercanas como las alejadas podrían plantear desafíos en el reconocimiento con redes neuronales.

Además de las distancias mencionadas, a 1, 2, 3 y 7 metros, se han empleado tres objetos de estudio, una silla, una bicicleta y un perro para enriquecer el análisis. La metodología de tratamiento y análisis de los datos se ha estructurado en dos componentes.



Figura 15: Diferentes ángulos de inclinación para un ángulo de orientación de 180°

En el primer enfoque se ha tratado la evaluación del ángulo de inclinación respecto IoU, cuyo objetivo es averiguar la diferencia que hay en los ángulos de inclinación 0°, 30°, 60° y 90°. En este caso, se mantiene constante el perfil 180° de posición que se puede observar en la Figura 15, y se realizan iteraciones entorno a la distancia y los ángulos de inclinación.



Figura 16: Comparación de las diferentes distancias en imagen RGB a 1,2,3 y 7 metros.

En segundo lugar, se ha realizado un análisis basado en la distancia del objeto, como se observa en la Figura 16. Para cada ángulo de inclinación, se han calculado las medias y desviaciones típicas correspondientes a las distintas distancias. Esta aproximación permite evaluar cómo las variaciones de la distancia influyen en la capacidad de reconocimiento de la red neuronal.

En el tercer enfoque de análisis, se han establecido de manera constante el ángulo de inclinación y como variable las 6 posiciones del suelo, como se puede observar en la Figura 17. Dentro de este marco, se ha llevado a cabo el cálculo de la media de las distancias y la desviación típica para cada una de estas posiciones, teniendo como variable el ángulo de posición. Con este enfoque, el propósito principal es investigar cómo la dirección del objeto puede influir en los resultados obtenidos.

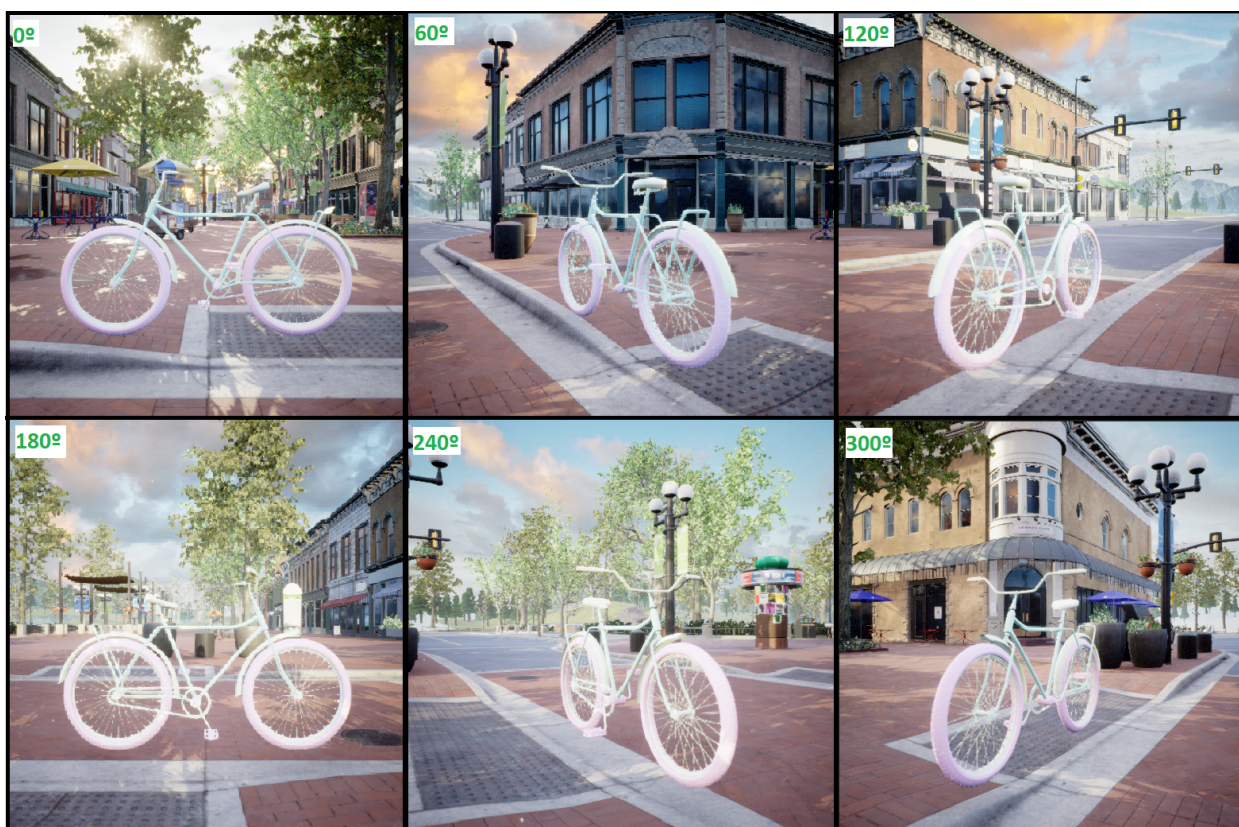


Figura 17: Diferentes ángulos de punto de vista para un ángulo de inclinación de 0°

5.2. Estudio sobre el grado de oclusión

Este experimento se centra en el análisis de la influencia de la oclusión en la detección de objetos. El objetivo es entender cómo la detección se ve afectada cuando el objeto de interés está parcialmente cubierto, muy cubierto y sin cubrir, como se observa en la Figura 18. Para abordar este objetivo, se ha incorporado una variedad de elementos que crean oclusión y además se ha fijado el ángulo de inclinación en 0° para su mejor estudio.

En consonancia con este enfoque, se ha aplicado la misma lógica de exploración a diferentes distancias de 2, 4 y 6 metros. Estas distancias representan distintas perspectivas desde las cuales se puede experimentar la oclusión del objeto. Además, para enriquecer el análisis, se han empleado tres objetos distintos: caballos, bicicletas y sillas.

Mediante este enfoque, se busca comprender cómo la oclusión, en combinación con la distancia y la naturaleza del objeto, impacta en la precisión de la detección. Los resultados de este estudio contribuyen a un entendimiento más profundo de cómo las redes neuronales responden ante condiciones de visibilidad limitada.



Figura 18: Comparación de las diferentes grado de oclusión en RGB: A la izquierda alta oclusión, a la derecha sin ocluir, en el centro parcialmente ocluido.

5.3 Estudio de complejidad en diferentes escenarios:

El objetivo de este experimento es analizar cómo el escenario influye en el reconocimiento de objetos. Para ello, se ha optado por seleccionar dos escenarios distintos: uno de ambiente urbano y otro de naturaleza, como se puede observar en la Figura 19. Además, el estudio involucra dos objetos de interés: una silla y una bicicleta. Las imágenes se han capturado desde dos distancias distintas, 3 y 4 metros.

Un aspecto interesante de este experimento es el cálculo de la media correspondiente a cada subcaso según el escenario en cuestión. Este enfoque permite obtener una visión más precisa de cómo el entorno específico impacta en el rendimiento del reconocimiento. La elección de ambientes diferentes, urbano y natural, añade profundidad al análisis al considerar cómo diferentes contextos influyen en la capacidad de las redes neuronales para reconocer los objetos seleccionados.



Figura 19: Escenario urbano y rural de la silla a 3 metros.

5.4 Estudio de la influencia de la posición no convencional del objeto de estudio.

El propósito de este experimento radica en investigar cómo la posición no convencional de un objeto afecta su reconocimiento. Para abordar este objetivo, hemos analizado situaciones donde los objetos se encuentran en disposiciones poco habituales. Los casos estudiados incluyen una silla tumbada, una bicicleta invertida y un caballo acostado, como se observa en la Figura 20. Siguiendo la continuidad de las investigaciones previas, estos experimentos se han llevado a cabo en el mismo escenario de entorno urbano y a distancias de 2, 4 y 6 metros.



Figura 20: Imagen de los 3 objetos en posición no convencional.

La elección de estos casos de estudio poco convencionales añade una dimensión interesante al análisis al considerar cómo la posición inusual puede influir en la capacidad de detección. La convergencia de estos elementos en el estudio permite arrojar luz sobre cómo la posición atípica de los objetos puede comprometer el rendimiento de las redes neuronales en la tarea de reconocimiento.

5.5 Estudio de complejidad de la escena

Para llevar a cabo este experimento, se ha introducido una variable adicional en el escenario: un barco grande que contrasta con la temática urbana previamente establecida, como se puede observar en la Figura 21. Es importante destacar que este nuevo objeto sigue perteneciendo a las 21 clases con las cuales la red neuronal ha sido entrenada. La particularidad de este ajuste radica en su capacidad para dificultar el proceso de reconocimiento y, por ende, en comprender su posible influencia en los resultados obtenidos.

La elección del objeto nuevo, un barco pesquero, se justifica por su tamaño significativamente mayor en comparación con los objetos en estudio (perro, caballo y silla). Esta disparidad de dimensiones busca intensificar el desafío de detección, contribuyendo a la exploración de cómo la red neuronal responde a situaciones de detección complejas.

En cuanto a las distancias utilizadas en el experimento, se ha trabajado con valores de 3, 4 y 5 metros. Estas distancias proporcionan una gama de perspectivas y condiciones para evaluar cómo la presencia del barco, junto con las distintas distancias, afecta el rendimiento de la detección.



Figura 21: Imagen del objeto de estudio caballo y en el fondo el objeto barco.

6. RESULTADOS

6.1 Estudio sobre la influencia del punto de vista

En el primer enfoque planteado, se ha realizado la media de las diferentes distancias (1,2,3 y 7 metros) para cada ángulo de inclinación fijando como constante el ángulo de posición 0° . Como se observa en la Figura 22, se puede considerar que la visión en los ángulos de 60° y 90° tienen valor cero en la totalidad de los resultados. Esto es probablemente debido a que la red neuronal no se ha entrenado para el reconocimiento de los objetos de estudio con un ángulo de inclinación elevado, sino que se ha entrenado para reconocer objetos en una visión frontal o ligeramente inclinada. Como se puede observar en la Figura 23, en la posición de 60° y 90° es nulo, en cambio en la posición 0° y 30° no ha llegado a afectar el ángulo de inclinación, reafirmando así la teoría del entrenamiento de la red.

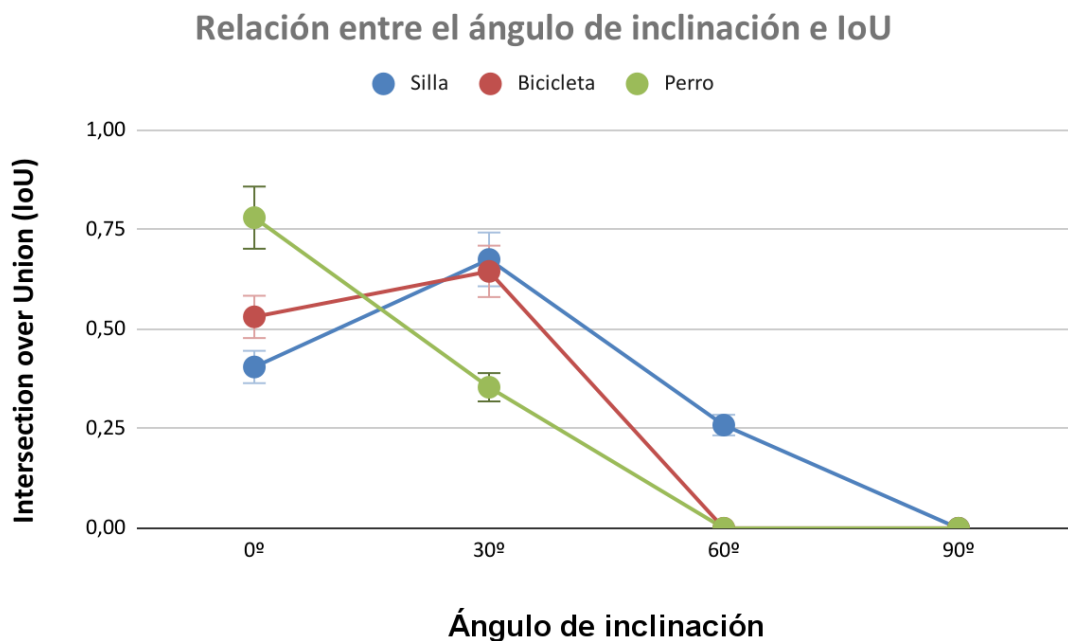


Figura 22: Relación entre IoU y el ángulo de inclinación para un ángulo de posición constante. Media y desviación típica para la distancia y punto de vista.



Figura 23: Imágenes RGB, GT y semántica para valores de ángulo de inclinación 0°, 30°, 60° y 90°.

En el segundo enfoque, se llega a la conclusión de que tanto la cercanía como la lejanía con respecto a la distancia en la que se encuentra el objeto de estudio tienen un impacto notable como se observa en la Figura 24, concretamente las distancias 2 y 3 metros tienen valores superiores a las 1 y 7 metros. Para el cálculo de la gráfica se han realizado las medias según el ángulo de inclinación para obtener resultados más significativos. Una apreciación sobre el cálculo de la media es que sus valores son ligeramente inferiores debido a que los valores de ángulo de inclinación con valores 60° y 90° son cercanos a cero.

Una información relevante acerca de esta gráfica es que la bicicleta muestra la menor afectación en términos generales. Esta particularidad podría ser atribuida al hecho de que el entrenamiento recibido por la red neuronal ha sido más completo en comparación con los otros dos objetos analizados.

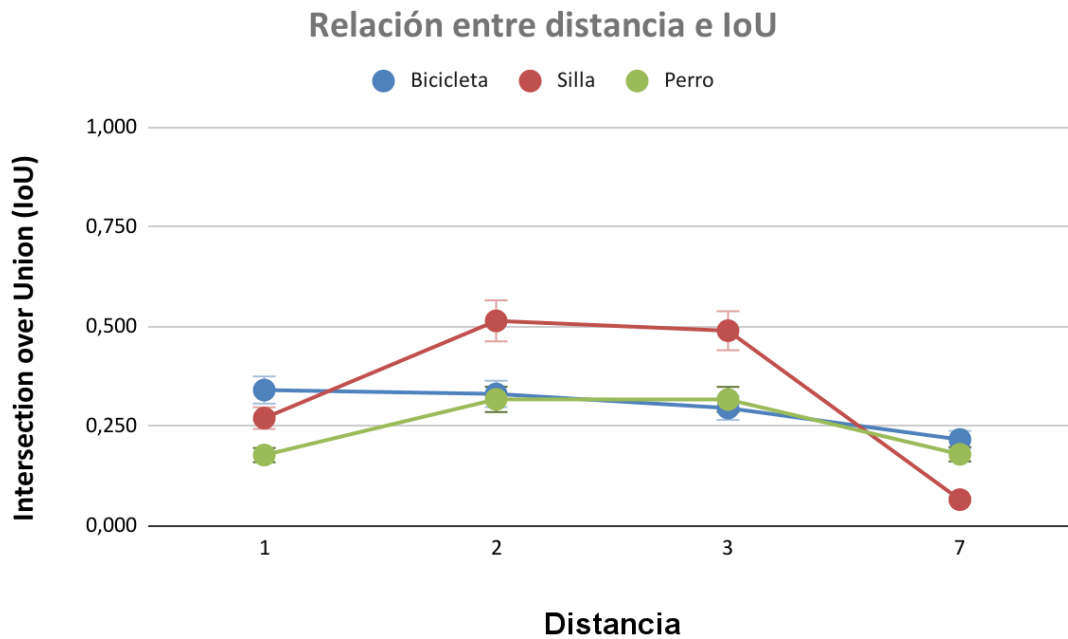


Figura 24: IoU frente a la distancia para diferentes clases. Media y desviación típica para el ángulo de inclinación y punto de vista.

Otro factor crucial se encuentra en la orientación del objeto. Se ha observado que la silla obtiene un reconocimiento significativamente más preciso cuando se encuentra en una posición frontal como se observa en la Figura 25 (ángulos 300º, 0º y 60º en la Figura 27), en contraste con situaciones en las que lo que se observa es su parte posterior, como en la Figura 26. Este dato resalta la importancia de la disposición del objeto en relación con la cámara y su impacto en la eficacia de la detección.

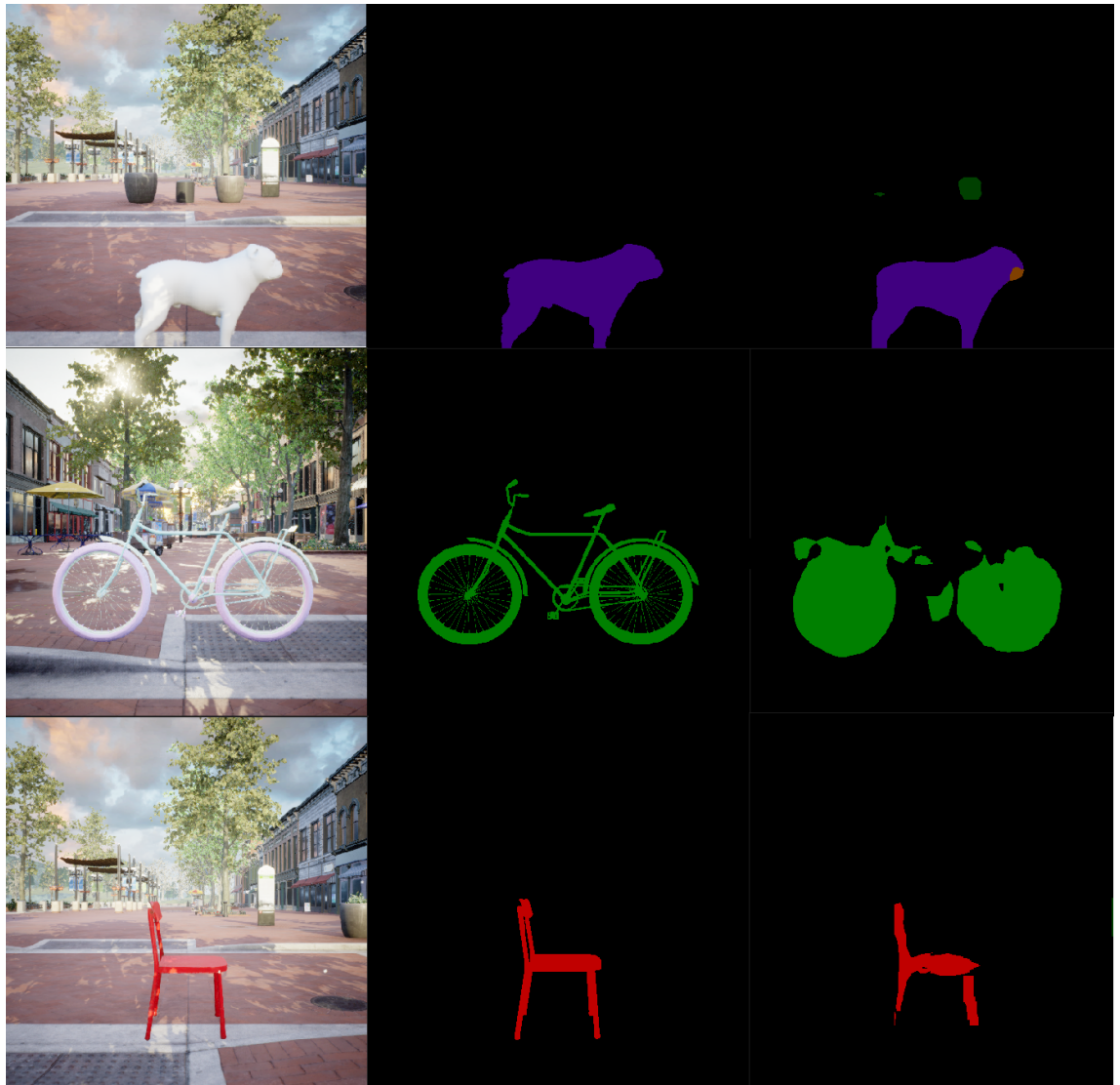


Figura 25: Posiciones de estudio para resultados de IoU altos. Para el perro 180º de punto de vista, 0º para la bicicleta y 180º para la silla.

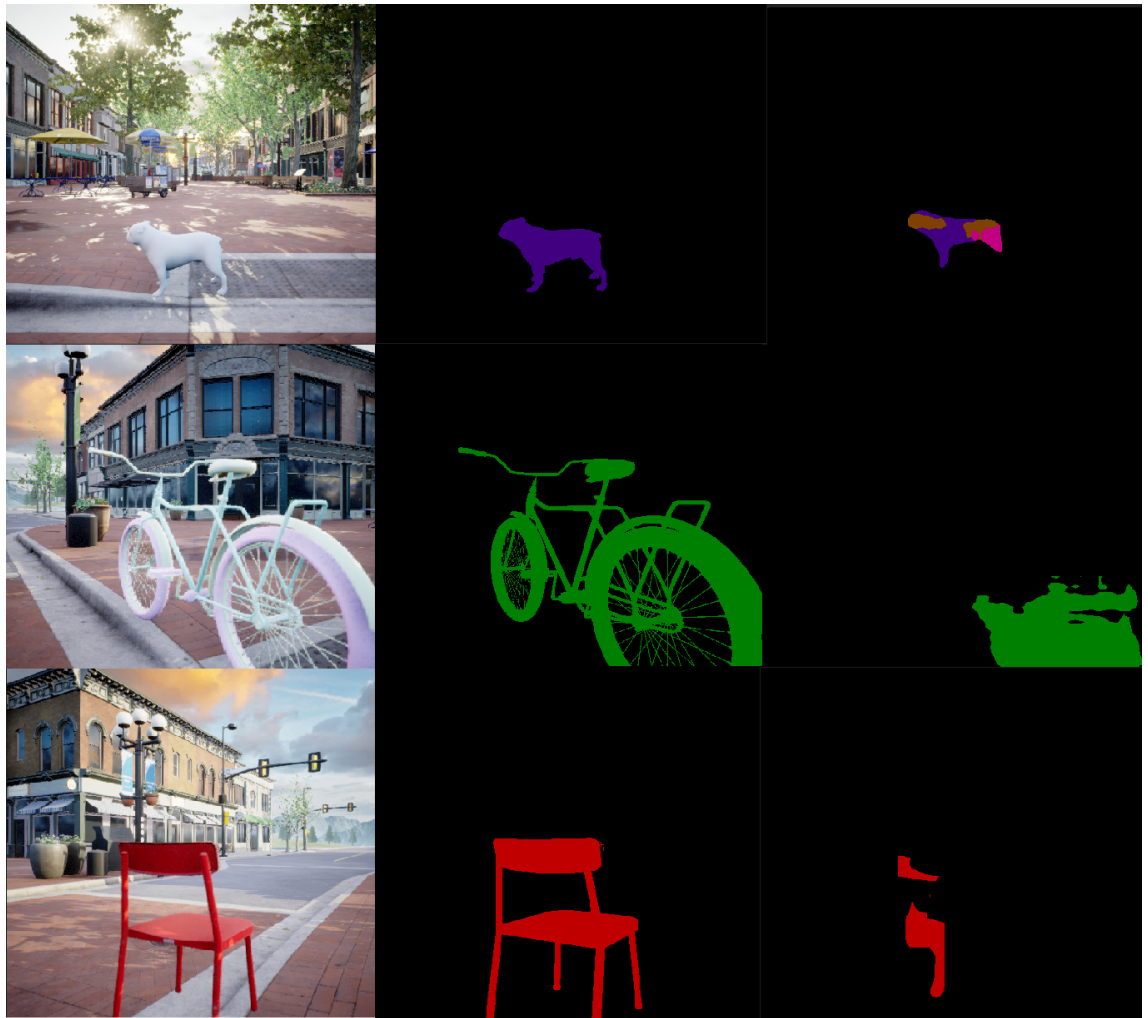


Figura 26: Posiciones de estudio para resultados de IoU bajos. Para el perro 0º de punto de vista, 60º para la bicicleta y 120º para la silla.

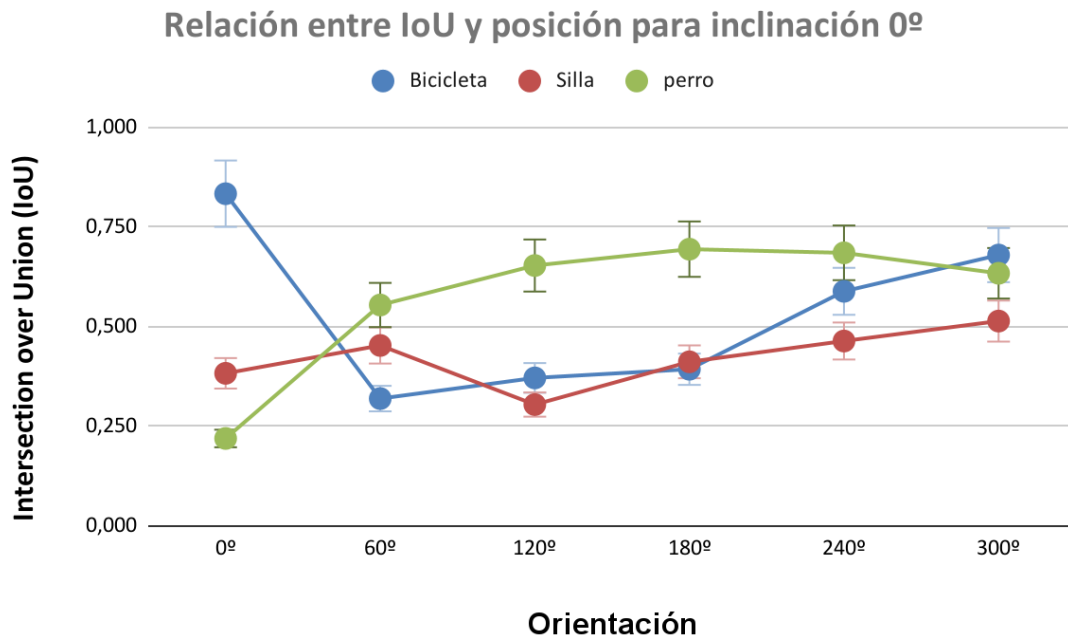


Figura 27: IoU frente a la orientación del punto de vista para 0º ángulo de inclinación constante y media sobre las distancias.

6.2 Estudio sobre el grado de oclusión

En líneas generales, se observa que la oclusión parcial tiene un impacto significativo en las vistas, como era previsible cuando un objeto se encuentra parcialmente cubierto. Los resultados más destacados se han obtenido en el caso del caballo, como se observa en la Figura 26, lo cual se puede inferir debido a su mayor tamaño en comparación con la bicicleta y la silla.

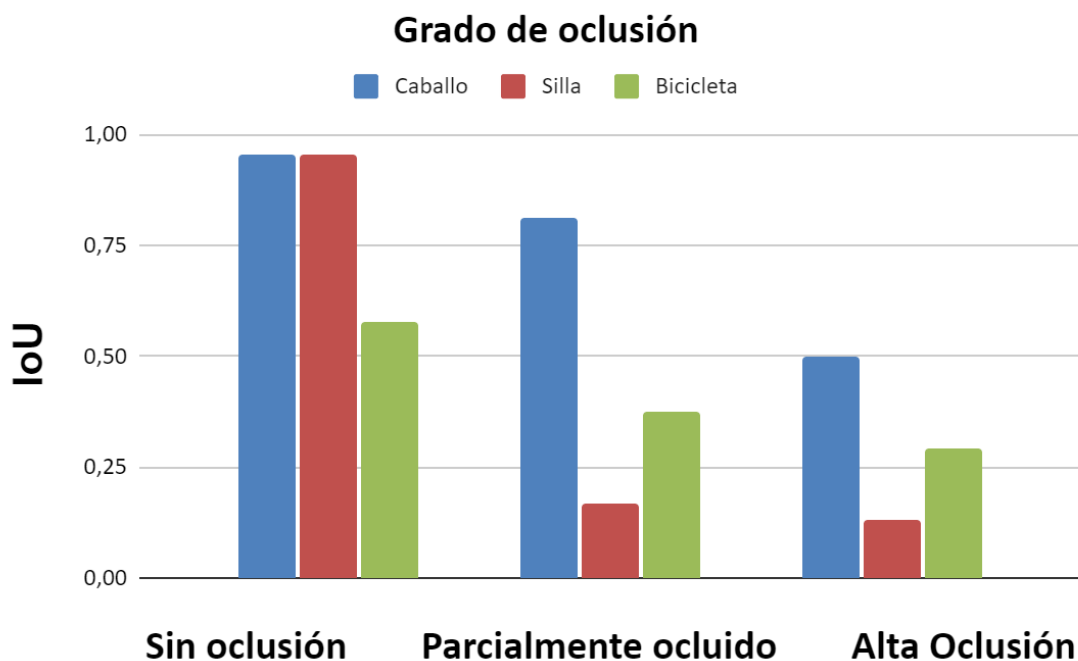


Figura 28 : Grado de oclusión para 0º ángulo de inclinación según el objeto de estudio.

En el caso de la silla, como se observa en la Figura 26, ha tenido los peores resultados siendo algunos incluso nulos esto es debido a que los obstáculos ocultan un mayor porcentaje de objeto lo que hace que dificulte su reconocimiento a la red neuronal. En la Figura 27 se puede apreciar la relevancia que tiene en el objeto de estudio mencionado.

La combinación de estos resultados sugiere la interacción compleja entre la oclusión, el tamaño del objeto y la distancia en la capacidad de las redes neuronales para realizar detecciones precisas. Estos hallazgos ofrecen información interesante sobre cómo estos factores contribuyen al rendimiento general de la detección en situaciones de oclusión parcial.

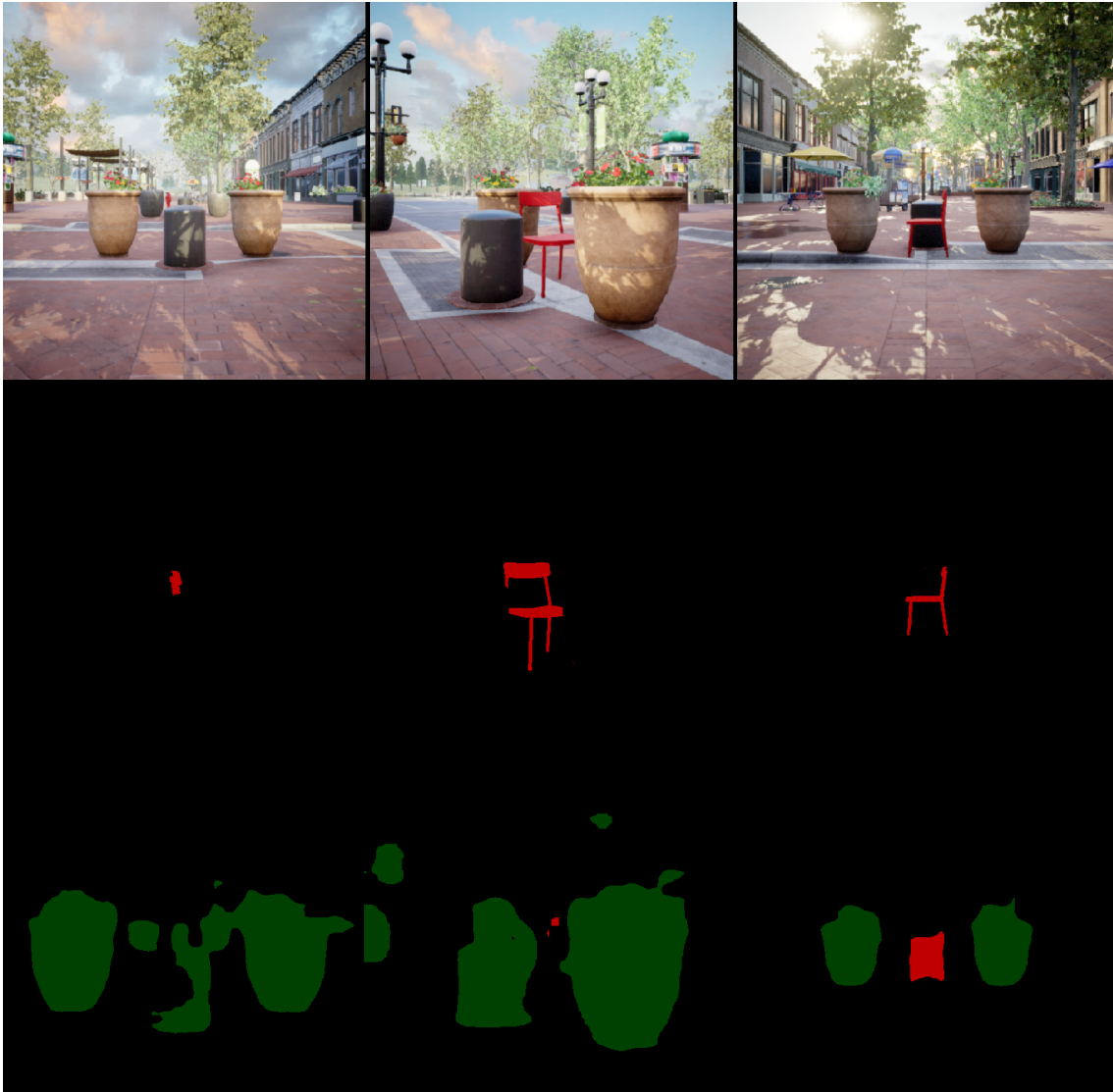


Figura 29 : Comparación de los diferentes grados de occlusión en RGB, GT y procesada: A la izquierda alta occlusión, a la derecha sin ocluir, en el centro parcialmente ocluido.

6.3 Estudio de los objetos en diferentes escenarios

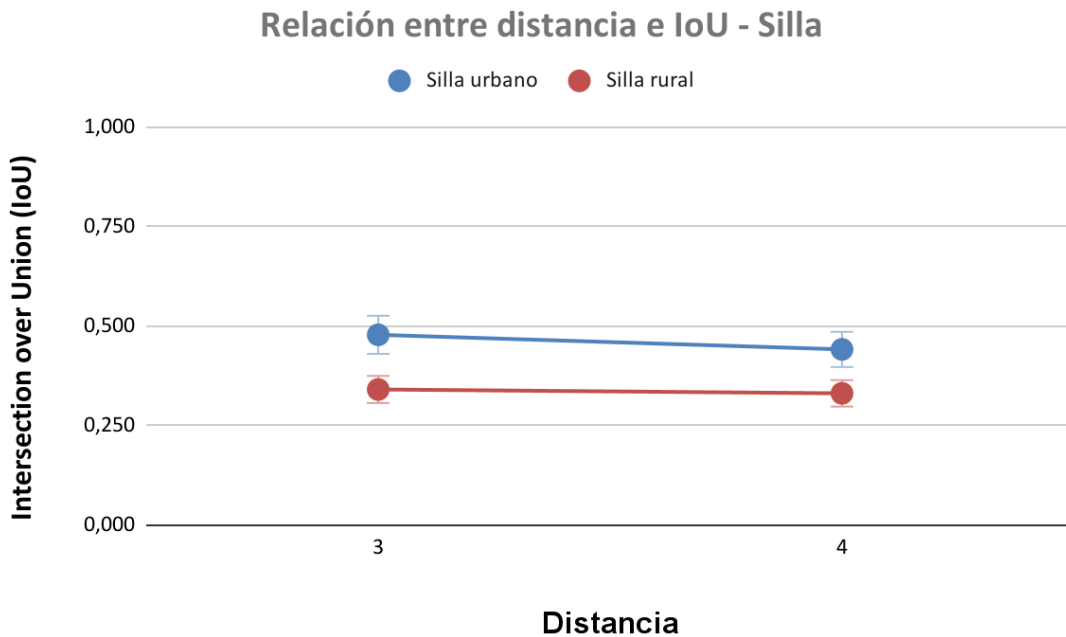


Figura 30 : Relación entre la distancia e IoU para Silla. Media y desviación típica para la inclinación y punto de vista.

Al analizar las medias, se extrae la conclusión de que la red neuronal demuestra un mejor desempeño en un entorno urbano en comparación con un entorno rural, como podemos observar en Figura 28 y Figura 29, respectivamente. Destaca en los datos que la métrica IoU registra valores ligeramente superiores en el caso de la silla. Esto puede interpretarse como una indicación de que las imágenes experimentales se asemejan más a las utilizadas previamente en el entrenamiento de la red neuronal.

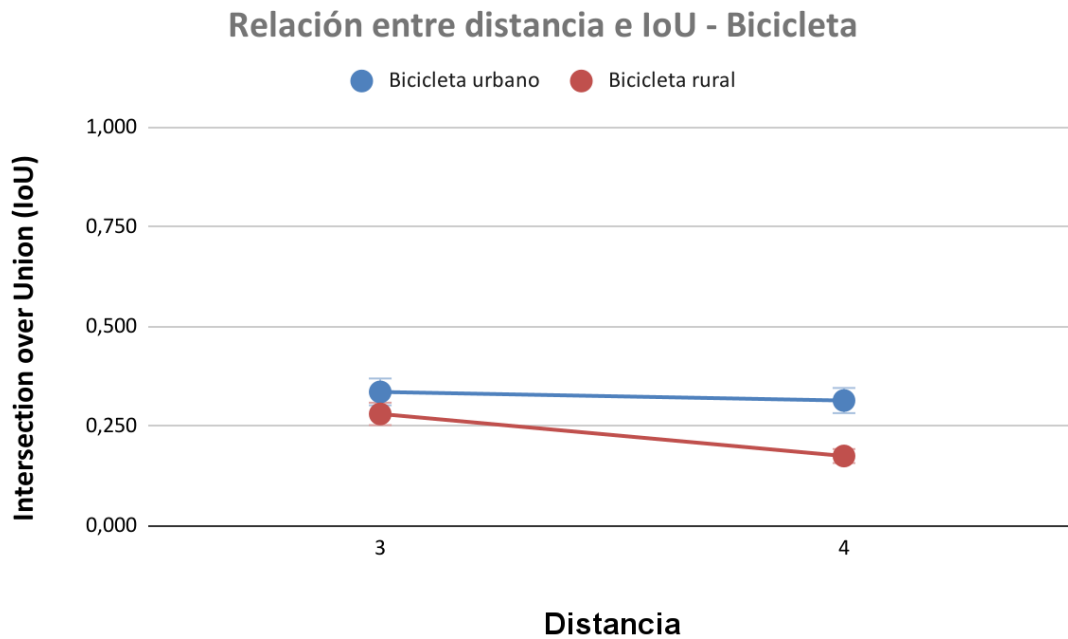


Figura 31 : Relación entre la distancia e IoU para Bicicleta. Media y desviación típica para la inclinación y punto de vista.

Siguiendo una línea similar a los estudios anteriores, se observa que la distancia tiene un efecto negativo en los resultados de las métricas IoU. Este hallazgo respalda la noción de que a medida que la distancia aumenta, la precisión de las predicciones se ve comprometida.

6.4 Estudio de la influencia de la posición no convencional del objeto

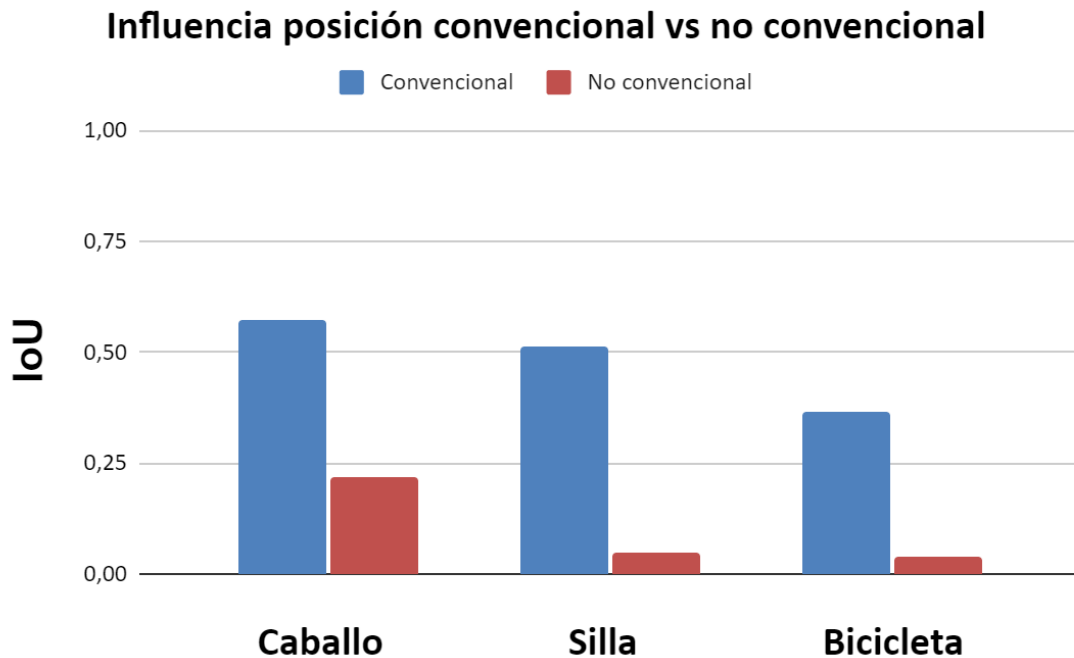


Figura 32 : Relación entre la posición convencional y no convencional

En las conclusiones de los tres experimentos realizados, surge un punto en común que merece ser resaltado: cuando un objeto se encuentra en una posición no convencional, su reconocimiento se vuelve considerablemente complicado, como se puede ver en la Figura 30. Se ha observado un caso excepcional en el contexto del caballo, donde se obtuvieron valores notoriamente altos de IoU para la orientación de 90º, cómo se puede apreciar en la Figura 31. Esto se debe a que, en dicha posición, el caballo parece estar en posición vertical, a pesar de que en la realidad está acostado. En el resto de los casos, los valores de IoU han sido notablemente bajos, incluso alcanzando valores de métrica de 0 en varios casos. Este patrón subraya la dificultad que presenta el reconocimiento en situaciones no convencionales.

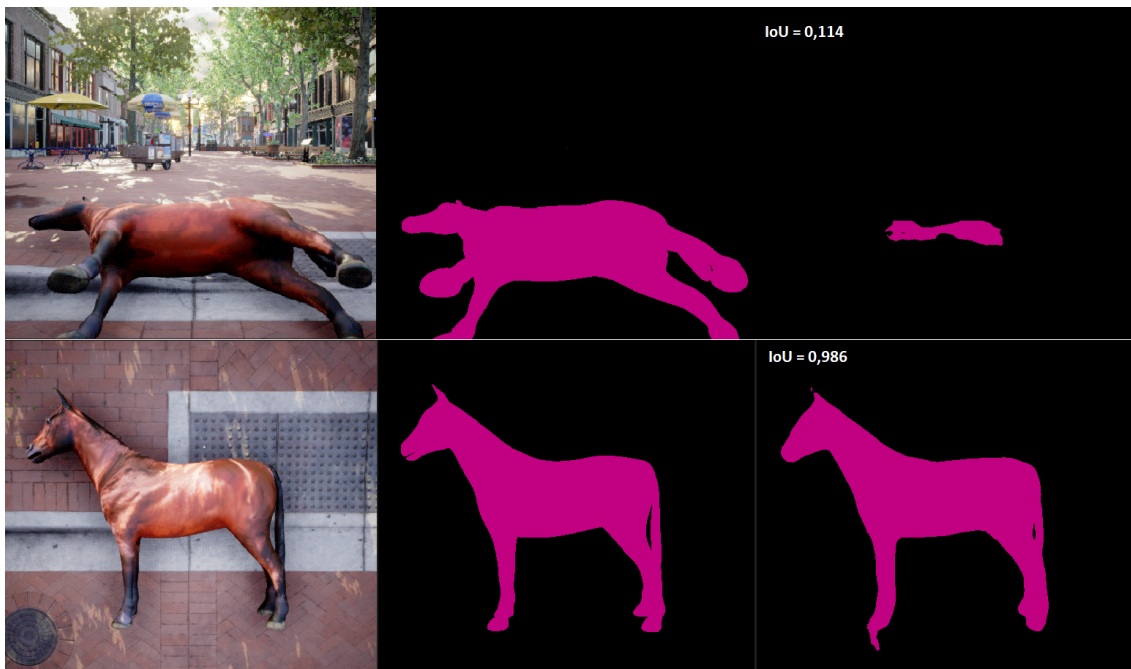


Figura 33 : Posición no convencional en 0º(arriba) y 90º (abajo) grados de inclinación para objeto de estudio caballo.

En los otros experimentos realizados con los objetos de estudio silla y bicicleta, los resultados han sido prácticamente nulo a excepciones como en la Figura 32, que la red neuronal ha llegado a detectar una rueda de la bicicleta.

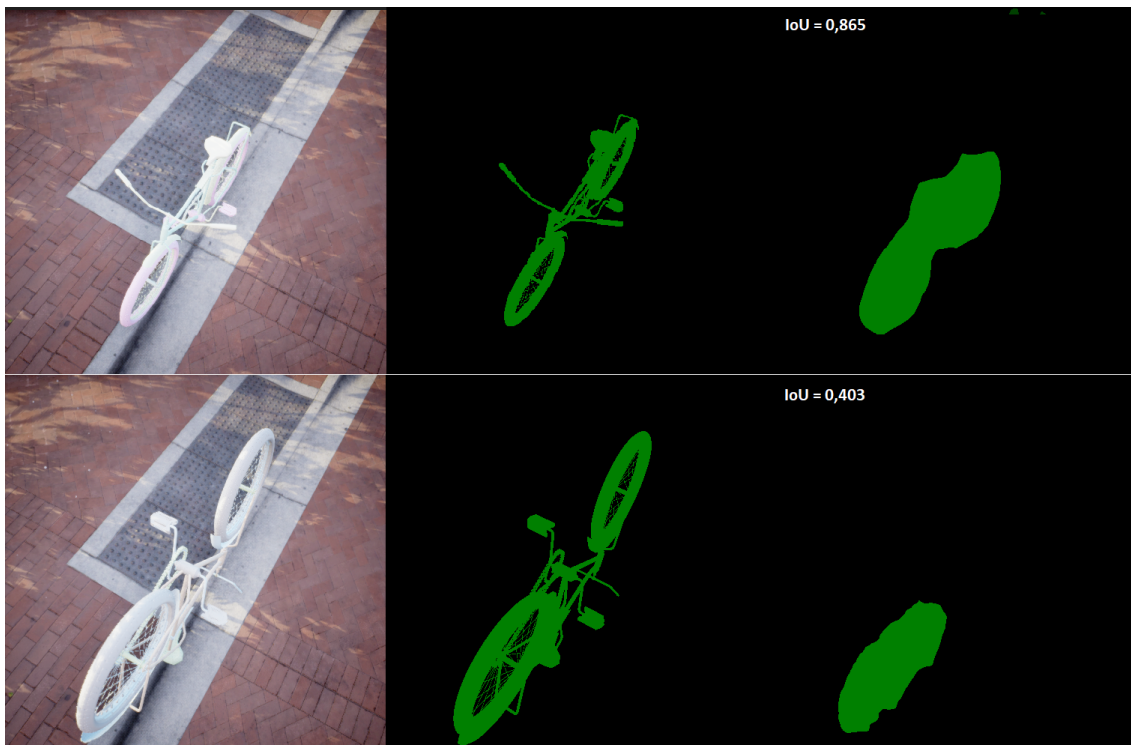


Figura 34 : Posición convencional (arriba) y no convencional (abajo) para objeto de estudio bicicleta.

Este hallazgo reviste importancia para el entrenamiento futuro de redes neuronales. Destaca la necesidad de incluir en el proceso de formación ejemplos de objetos en posiciones poco habituales, para dotar a las redes neuronales de una capacidad más sólida de adaptación y generalización.

6.5 Estudio de complejidad de la escena

El enfoque de este estudio fue incrementar el nivel de complejidad en la detección de objetos al introducir un barco en el escenario. Sin embargo, sorprendentemente, esta adición no generó un impacto sustancial en el rendimiento de la red neuronal. A pesar de la inclusión del barco, los valores de detección se mantuvieron en niveles comparables a los obtenidos en escenarios con fondos diferentes, como se observa en la Figura 33.

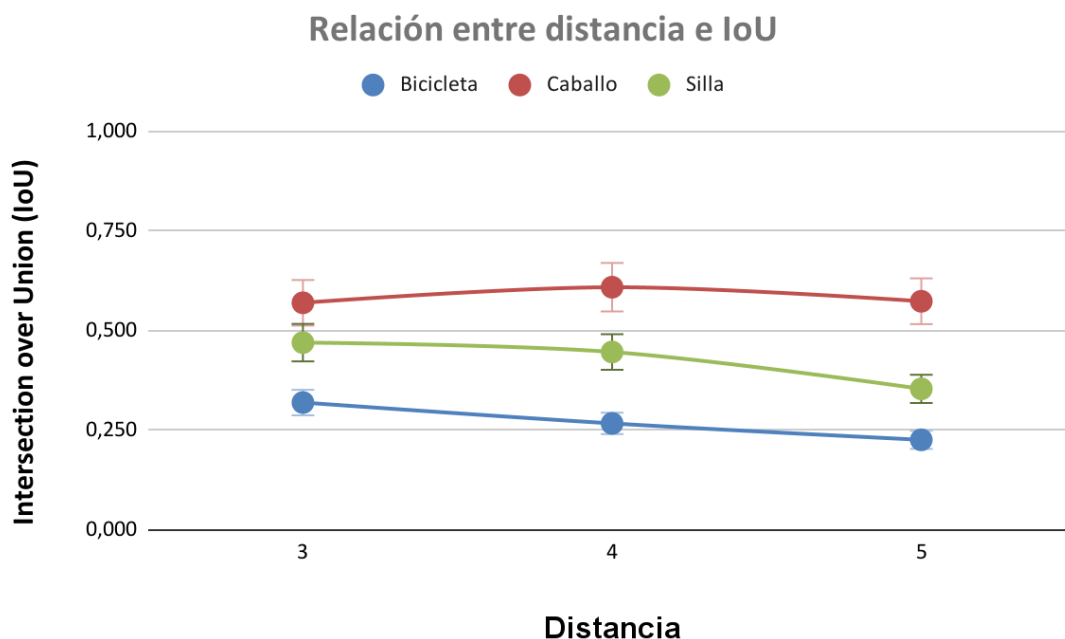


Figura 35 : Relación entre la distancia e IoU para cada objeto. Media y desviación típica para la inclinación y punto de vista.

Este resultado refleja la capacidad de la red neuronal para generalizar y adaptarse a nuevas situaciones, incluso cuando se presentan objetos inesperados en el entorno, como se puede observar en la Figura 34. Así, se resalta la robustez y versatilidad de la red en términos de reconocimiento de objetos en contextos diversos.

En línea con las tendencias experimentales, se constata nuevamente que la distancia juega un papel negativo en la precisión de los resultados. Este hallazgo es consistente con los experimentos anteriores y subraya la importancia de considerar las distancias en el diseño y análisis de futuros estudios.

En cuanto a los ángulos de inclinación, se reafirma la eficacia de las perspectivas frontales (0º) y ligeramente inclinadas (30º) en términos de obtener los mejores resultados. Estos ángulos parecen brindar a la red neuronal una visión óptima para el reconocimiento preciso de los objetos.

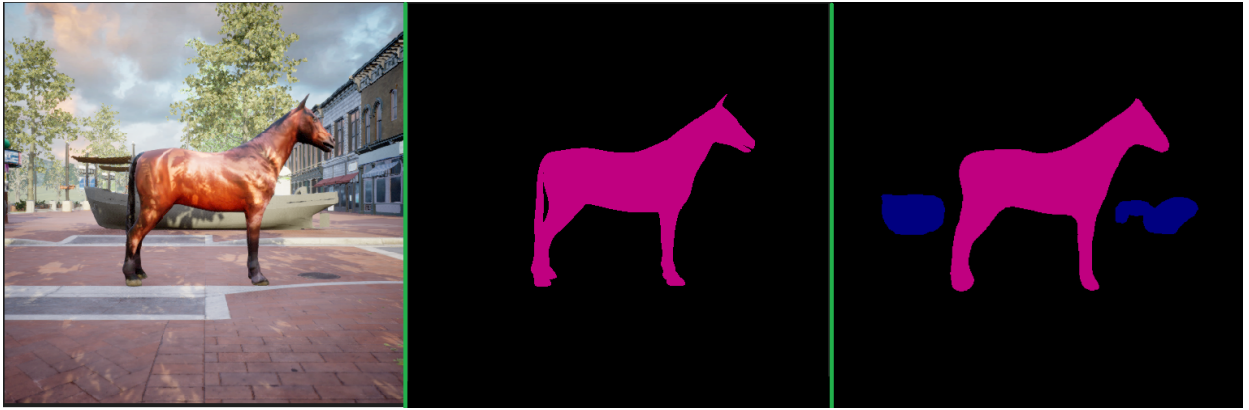


Figura 36: Imagen RGB, imagen GT y imagen semántica procesada por red neuronal.

7. CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURAS

En este trabajo fin de grado se han adquirido conocimientos para la implementación de entornos virtuales fotorrealistas con el objetivo de estudiarla influencia de la posición de cámara en el problema de reconocimiento de objetos utilizando redes neuronales. Una vez se dispone de los entornos se ha procedido al diseño de experimentos generando diversas situaciones que generan problemas comunes en redes neuronales. Los estudios muestran que estas situaciones generan en efecto problemas como la distancia al objeto de estudio, ángulo de inclinación de inferencia o entorno de trabajo pero que estos se pueden solucionar mediante la correcta elección de puntos de vista o configuraciones de cámara. El entorno desarrollado permite el estudio de redes neuronales "out-of-the-shelf" y su correcta caracterización para el uso en entornos reales. Esto supone un avance con respecto a los métodos actuales que únicamente muestran métricas de calidad utilizando imágenes de datasets de validación sin relación entre ellas y sin tener en cuenta la influencia de la perspectiva y otros factores externos. Utilizando el método de evaluación propuesto en el TFG para el estudio de perspectiva y factores externos se puede evaluar de forma más cercana a la realidad la capacidad de las redes entrenadas.

Tras las conclusiones obtenidas en el Trabajo Fin de Grado es posible afrontar algunas líneas futuras, con el fin de mejorar los resultados obtenidos. Una opción viable para mejorar el rendimiento es considerar la posibilidad de reentrenar la red en situaciones donde se ha observado un menor desempeño, como en los casos de 60º y 90º. Esta dirección exploratoria busca analizar cómo la red neuronal respondería a un proceso

de reentrenamiento y qué implicaciones acarrearía dicho procedimiento en su rendimiento general. Una línea de investigación adicional podría dirigirse hacia la modificación de la red neuronal convolucional Resnet 50 seleccionada, optando por una arquitectura más complejas o eficientes. Esta comparación entre arquitecturas podría arrojar luz sobre la adaptabilidad y pertinencia de distintas estructuras en condiciones variadas de posicionamiento. Estas posibles direcciones futuras se erigen como oportunidades para medir la relación entre el coste computacional y las ventajas reales de complejidad frente a situaciones adversas.

REFERENCIAS

- [1] Instituto Immune. (2022). Librerías Python que son. Immune Institute. Recuperado de: <https://immune.institute/blog/librerias-python-que-son/>
- [2] JetBrains. (2023). Pycharm. Recuperado de: <https://www.jetbrains.com/pycharm/>
- [3] "Unreal Engine." Epic Games. <https://www.unrealengine.com/>
- [4] "AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles." Microsoft Research. <https://www.microsoft.com/en-us/research/project/airsim/>
- [5] PyTorch. (2023). DeepLabV3 ResNet-50. Recuperado de: https://pytorch.org/hub/pytorch_vision_deeplabv3_resnet50/
- [6] Varela Arregocés, E., & Campbells Sánchez, E. (Junio 2011). Redes Neuronales Artificiales: Una Revisión del Estado del Arte, Aplicaciones Y Tendencias Futuras.
- [7] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444
- [8] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3431–3440, 2015.
- [9] S. C. Yurtkulu, Y. H. Şahin and G. Unal, "Semantic Segmentation with Extended DeepLabv3 Architecture," 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 2019, pp. 1-4, doi: 10.1109/SIU.2019.8806244.
- [10] A. Rosebrock, "Intersection over union (iou) for object detection," Jun 2016.
- [11] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 660-662.
- [12] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. *En Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 658-666).
- [13] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Disponible en: <https://www.deeplearningbook.org/>
- [14] Disponible en: <https://newsroom.porsche.com/es/2021/tecnologia/PLA-porsche-engineering-tecnologia-conduccion-autonoma-nvidia-24461.html>
- [15] Disponible en: https://es.wikipedia.org/wiki/Perceptr%C3%B3n_multicapa

[16] Disponible en: https://rpubs.com/luis_castillo/redes_neuronales

[17] **Disponible en:**

http://reaxion.utleon.edu.mx/Art_Redес_neuronales_para_la_toma_de_decisiones_en_el_sector_agr%C3%ADcola_analisis_exploratorio.html

[18] **Disponible en:** <https://www.azorobotics.com/Article.aspx?ArticleID=553>

[19] **Disponible en:**

<http://pompeisites.org/en/comunicati/spot-a-quadruped-robot-at-the-service-of-archaeology-to-inspect-archaeological-areas-and-structures-in-safety/>

[20] **Disponible en:** <https://cnvrg.io/semantic-segmentation/>

[21] **Disponible en:** <https://aprendeconalf.es/docencia/python>

[22] **Disponible en:** <https://aprendepython.es/>

[23] **Disponible en:** <https://code.visualstudio.com/docs/getstarted/settings>