



ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
INFORMÁTICOS

UNIVERSIDAD POLÍTÉCNICA DE MADRID

Reconocimiento de landmarks visuales basado en Deep Learning + CNNs para la navegación autónoma en entornos mixtos de interior/exterior mediante mapas topológicos visuales

TRABAJO FIN DE MÁSTER
MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL

AUTOR: Amable José Valdés Cuervo
TUTOR: Darío Maravall Gómez-Allende

Agradecimientos

TODO Extensión máxima de una página

Resumen

TODO Realizar el resumen al finalizar de la documentación

Abstract

TODO Realizar el abstract al finalizar la documentación

Índice general

Agradecimientos	III
Resumen	V
Abstract	VII
1. Introducción	1
1.1. Antecedentes	1
1.2. Objetivos	2
2. Descripción del entorno experimental de navegación	5
2.1. Selección del entorno	5
2.2. Mapa topológico	6
2.3. Landmarks seleccionados	7
2.4. Navegación	13
3. Generación del dataset de imágenes de landmarks para el entrenamiento y evaluación de los reconocedores	15
3.1. Creación del dataset	15
3.1.1. Toma de imágenes	15
3.1.2. Preprocesamiento	15
3.2. Contenido del Dataset	15
4. Clasificador: Deep Learning + CNNs	19
4.1. Redes Neuronales	19
4.2. CNNs - Convolutional Neural Networks	19
4.3. Técnicas usadas	19
5. Validation: test estáticos	21
6. Test: tests dinámicos	23
7. Conclusiones y líneas futuras de trabajo	25
7.1. Conclusiones	25
7.2. Otras líneas de trabajo	25

A. Anexos	27
Bibliografía	29

Índice de figuras

1.1.	Mapa estudiado en la asignatura de robots autónomos	1
2.1.	Mapa de la zona de estudio	5
2.2.	Mapa topológico del proyecto	6
2.3.	Áreas de estudio de cada nodo	7
2.4.	Ejemplos de imágenes pertenecientes al nodo 1, el pasillo de las habitaciones	8
2.5.	Ejemplos de imágenes pertenecientes al nodo 2, las escaleras del hall .	9
2.6.	Ejemplos de imágenes pertenecientes al nodo 3, el comedor	10
2.7.	Ejemplos de imágenes pertenecientes al nodo 4, el edificio de las habitaciones	10
2.8.	Ejemplos de imágenes pertenecientes al nodo 5, el parking	11
2.9.	Ejemplos de imágenes pertenecientes al nodo 6, la biblioteca	12
2.10.	Ejemplos de imágenes pertenecientes al nodo 7, el gimnasio	12
3.1.	Dataset para Redes Neuronales	16
3.2.	División del dataset para este proyecto	17

Índice de tablas

Índice de algoritmos

Capítulo 1

Introducción

1.1. Antecedentes

Parto de un trabajo anterior realizado en la asignatura de robots autónomos de este mismo máster en el que mis compañeros y yo realizamos un diseño de un mapa topológico y entrenamos un K-NN para identificar la zona donde se encontraba.

El mapa topológico que diseñamos representaba el segundo piso del bloque 2-3; el camino desde las escaleras del bloque 2 hasta el aula 3202. Dicho mapa topológico puede verse en la Figura 1.1.

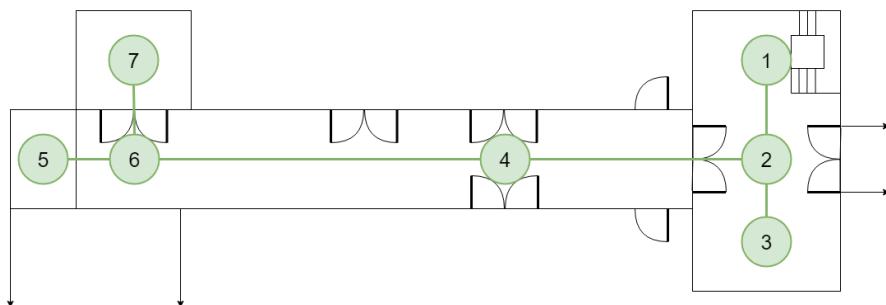


Figura 1.1: Mapa estudiado en la asignatura de robots autónomos

Cada nodo representado en el mapa topológico de la Figura 1.1 es un landmark visual representativo de cada zona del pasillo y los arcos indican desde qué nodo se puede viajar hasta cuales nodos.

El dataset: Entrenamos al K-NN con un dataset que constaba de unas 4200 imágenes, de las que 4090 fueron usadas para entrenamiento y 110 para validación, y un vídeo para test. Esto se traduce en que para cada uno de estos nodos se tomaron unas 600 imágenes de las que alrededor de unas 590 sirvieron para entrenamiento y alrededor de unas 10 para validación.

Para usar estas imágenes para entrenar al K-NN convertimos esas imágenes en color a imágenes en escala de grises y, a su vez, esas imágenes en escala de grises a

histogramas. Mediante el estudio de estos histogramas el K-NN podía predecir, dada una nueva foto que nunca hubiera visto, cuál era la imagen que más se le parecía de entre todas las imágenes de entrenamiento.

- Dataset, 4200 imágenes y 1 vídeo.
 - Training, 4090 histogramas de imágenes en escala de grises.
 - Validation, 110 histogramas de imágenes en escala de grises.
 - Test, histogramas de los frames de 1 único vídeo con el recorrido a estudiar.

Conclusiones: Los resultados de esta aproximación fueron los esperados; buenos resultados en validación y malos resultados en test.

- En validación el K-NN tuvo una tasa de aciertos del 99 %, acertando 109 de las 110 imágenes.
- En test el K-NN tuvo una tasa de aciertos del 46 %, acertando 552 frames de los 1192 frames del vídeo.

Con un K-NN, que es un clasificador sencillo, los resultados en la validación son muy buenos, pero es un clasificador que no generaliza ya que simplemente se fija en los niveles del histograma de la imagen y debido a ello los resultados en test son muy malos como se podía esperar, no llegando a acertar el nodo en el que se encuentra ni la mitad de las veces. Además, hay que añadir que el dataset es pequeño; con unas 600 imágenes por nodo y tan solo 10 imágenes para la validación es normal que los resultados no sean de mejor calidad.

1.2. Objetivos

Con este proyecto espero conseguir una mejora significativa de los resultados obtenidos en el proyecto pasado usando técnicas novedosas en el campo de la visión por computador como es el Deep Learning junto al uso de redes convolucionales o CNNs.

Para conseguir esto se van ha de realizar una serie de tareas:

- **Nuevo mapa topológico:** Se diseñará un nuevo mapa topológico con entornos mixtos de interior y exterior con landmarks bien diferenciados.
- **Generación de un nuevo dataset:** Se tomarán imágenes de los landmark visuales característicos del mapa topológico.
- **Deep Learning:** Se modelizará una red neuronal que será entrenada con el nuevo dataset.

- **Estudio de los resultados:** Se le realizarán pruebas a la red neuronal con los datos de validación y test. Se estudiarán estos resultados centrándose en la exactitud de sus predicciones y sus errores, comprobando así la eficiencia y rendimiento de la red neuronal.

Capítulo 2

Descripción del entorno experimental de navegación

2.1. Selección del entorno

En este proyecto voy a trabajar sobre un entorno distinto al presentado en los antecedentes en 1.1. El anterior entorno era completamente de interior y he decidido ampliarlo a un entorno mixto de interior/exterior, siendo este nuevo entorno los terrenos de la residencia de estudiantes donde he residido durante el periodo lectivo. Dicho entorno puede verse representado en la Figura 3.2.



Figura 2.1: Mapa de la zona de estudio

Las razones de esta ampliación son varias:

- **Texturas y formas:** En el entorno de los antecedentes prevalecen solo un par de texturas y/o colores: el suelo, negro o beige; las paredes, siempre blancas y con una franja gris o negra; algún marrón para elementos de madera como

las puertas. Con este nuevo entorno tenemos mucha más variedad de texturas: paredes pintadas de distintos colores; suelos de ladrillo, piedra, madera o distintos colores; edificios con texturas características.

Esto dificulta el entrenamiento de la red, ya que esta ha de generalizar muchas más texturas y elementos, pero da mucho juego para la identificación de los landmark y enriquece el entorno de experimentación.

- **Iluminación:** Al trabajar en un entorno mixto de interiores/exteriores la iluminación en muchos landmark será distinta dependiendo de las condiciones climáticas (sol, nubes, lluvia), no como ocurría en el entorno de los antecendentes donde solo nos preocupábamos por la iluminación para comparar los histogramas y donde no había tanta variedad.

Esto dará ventajas y desventajas a la hora de realizar el dataset: como desventaja, hay que asegurarse de tomar imágenes en varias horas del día con distinta iluminación y condiciones climáticas; como ventaja, la red generalizará mejor con imágenes con distinta iluminación, lo que nos dará mejores resultados.

2.2. Mapa topológico

He analizado el entorno experimental de la misma manera que se realiza en [2]; buscando lugares representativos que podrían servir como nodos clave para definir los landmarks visuales del mapa topológico y estudiando como se puede navegar desde un landmark a otro.

Tras el análisis inicial he llegado a la confección del mapa topológico que puede observarse en la Figura 2.2.



Figura 2.2: Mapa topológico del proyecto

En dicho mapa puede observarse la relación entre los distintos nodos y como desde un nodo se llega a otros a través de la conexión de sus arcos. Más adelante en el apartado 2.3 explicaré cada nodo en detalle.

2.3. Landmarks seleccionados

Concreción vs Generalidad

En [1] podemos ver como se seleccionan los landmark visuales mediante la selección de objetos concretos y bien diferenciados.

Habitualmente el enfoque que se toma es ese; se seleccionan objetos concretos y únicos del entorno (como puede ser en [2] una puerta o un extintor) y se enseña a un clasificador a identificar estos objetos. No obstante, en este proyecto se va a realizar un enfoque distinto.

El entorno de los antecedentes o de [1, 2] son entornos pequeños; una habitación o pasillos de unos metros de distancia como mucho. En el entorno que hemos definido en este proyecto tenemos media hectárea de terreno de estudio.

Debido al tamaño del entorno las imágenes que tomaremos para nuestro dataset serán imágenes muy genéricas que contendrán muchos objetos y estructuras que se repetirán a lo largo de todos los nodos tales como puertas, carteles, escaleras... Además de tener presente el hecho de que muchos de nuestros landmark son objetos compuestos por otros objetos, como pueden ser los edificios que clasificaremos que están definidos por ventanas, puertas, chimeneas, etc.

Este entorno podría dar problemas a clasificadores que se centran en buscar objetos concretos, y es por eso por lo que para este proyecto nos vamos a apoyar en redes neuronales y, en especial, en las redes convolucionales donde la generalización en imágenes es uno de los aspectos claves usados para la correcta clasificación.

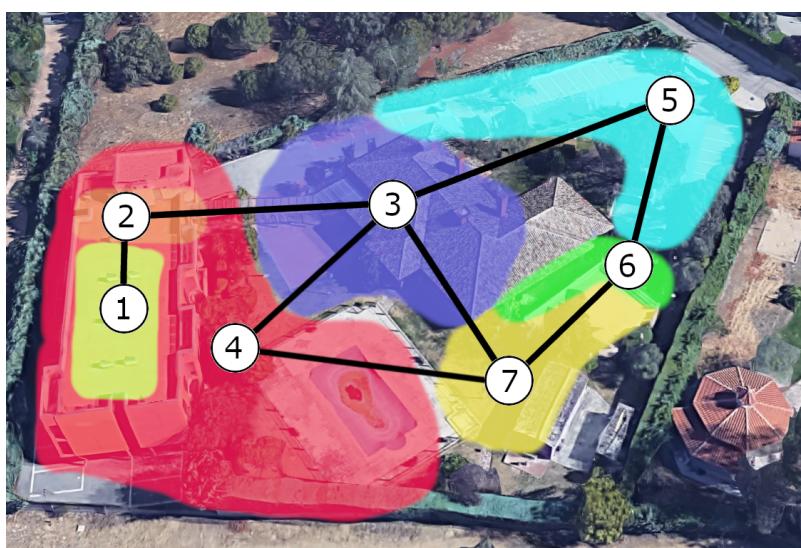


Figura 2.3: Áreas de estudio de cada nodo

De esta manera tomaremos imágenes desde distintos ángulos de los escenarios que queremos clasificar y dejaremos que la red neuronal encuentre patrones comunes en la imagen para poder llegar a una predicción acertada del lugar donde se encuentra. Para esto hemos definido una serie de zonas donde tomaremos imágenes de cada landmark. Estas áreas definidas se pueden ver en la Figura 2.2.

Las imágenes del dataset serán tomadas desde las áreas presentadas en la Figura 2.2 enfocadas en un landmark más genérico que puede estar compuesto por muchos objetos que incluso pueden o no estar en la imagen. Esto quiere decir que seguiremos teniendo la relación "1 landmark = 1 nodo", pero el landmark será más complejo.

En el siguiente apartado 2.3 explicaré en detalle cada uno de los nodos seleccionados y su landmark asociado, definiendo que objetos aparecerán en la imagen y cómo esperamos que la red generalice para poder identificar el nodo.

Landmarks en detalle

A continuación explicaré en detalle cada uno de los nodos definidos.

■ Nodo 1 - Pasillo de las habitaciones

- *Descripción:* Dentro del edificio donde se encuentran las habitaciones de los estudiantes este nodo comprende todo el pasillo del edificio.
- *Conexiones:* Desde este nodo se puede ir al nodo 2, las escaleras del hall.
- *Landmark:* El pasillo.
- *Generalización:* La red deberá generalizar usando: el pasillo, las paredes distintivas de color blanco y verde, las puertas de las habitaciones y las ventanas que dan al exterior.
- *Ejemplos:* Se pueden ver algunos ejemplos de este nodo en la Figura 2.4



Figura 2.4: Ejemplos de imágenes pertenecientes al nodo 1, el pasillo de las habitaciones

■ **Nodo 2 - Escaleras del hall**

- *Descripción:* Dentro del edificio donde se encuentran las habitaciones de los estudiantes este nodo comprende el hall del edificio.
- *Conexiones:* Desde este nodo se puede ir a: el nodo 1, el pasillo de las habitaciones; y el nodo 3, el comedor.
- *Landmark:* La escalera.
- *Generalización:* La red deberá generalizar usando: el hall, las escaleras del hall, las paredes distintivas de color blanco, la puerta del ascensor y el dispensador de agua.
- *Ejemplos:* Se pueden ver algunos ejemplos de este nodo en la Figura 2.5

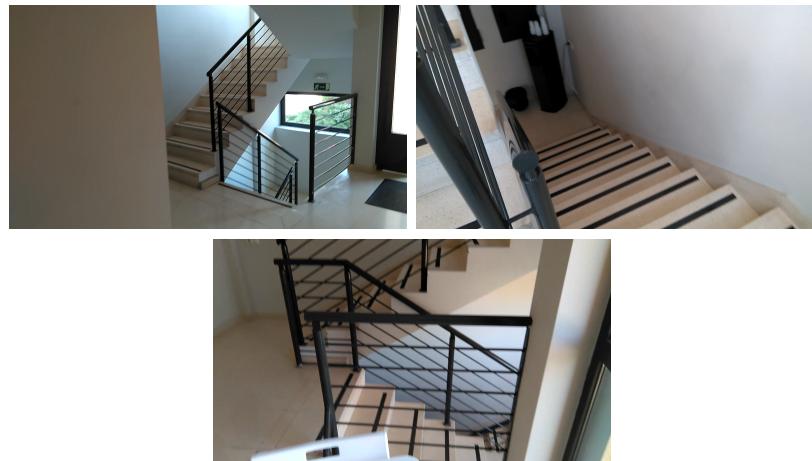


Figura 2.5: Ejemplos de imágenes pertenecientes al nodo 2, las escaleras del hall

■ **Nodo 3 - Comedor**

- *Descripción:* En el exterior se encuentra la fachada del edificio donde se encuentra el comedor y las cocinas de la residencia.
- *Conexiones:* Desde este nodo se puede ir a: el nodo 2, el hall; el nodo 4, la fachada del edificio de las habitaciones; el nodo 7, el gimnasio; y el nodo 5, el parking.
- *Landmark:* La fachada del comedor.
- *Generalización:* La red deberá generalizar usando: el edificio donde se encuentra el comedor y las cocinas, que a su vez se encuentra compuesto por ventanas, puertas y la chimenea; los cubos de basura; sillas y mesas de exterior; las texturas del suelo y los ladrillos del edificio; algunos arboles de color característico.
- *Ejemplos:* Se pueden ver algunos ejemplos de este nodo en la Figura 2.6

■ **Nodo 4 - Edificio de las habitaciones**



Figura 2.6: Ejemplos de imágenes pertenecientes al nodo 3, el comedor

- *Descripción:* En el exterior se encuentra la fachada del edificio donde se encuentran las habitaciones de los estudiantes.
- *Conexiones:* Desde este nodo se puede ir a: el nodo 3, el comedor; y el nodo 7, el gimnasio. El nodo 1 y 2 están dentro de este edificio.
- *Landmark:* La fachada de las habitaciones.
- *Generalización:* La red deberá generalizar usando la fachada del edificio donde se encuentran las habitaciones de los estudiantes, que a su vez se encuentra compuesto por ventanas, puertas, la chimenea, el cartel que indica que se trata del edificio de las habitaciones y las formas cuadradas o rectangulares pertenecientes a la arquitectura del edificio o los colores con los que ha sido pintada la fachada.
- *Ejemplos:* Se pueden ver algunos ejemplos de este nodo en la Figura 2.7



Figura 2.7: Ejemplos de imágenes pertenecientes al nodo 4, el edificio de las habitaciones

■ Nodo 5 - Parking

- *Descripción:* En el exterior se encuentra el parking desde el que se consigue acceso a la calle y al exterior de la residencia.
- *Conexiones:* Desde este nodo se puede ir a: el nodo 3, el comedor; y el nodo 6, la biblioteca.
- *Landmark:* La puerta del parking
- *Generalización:* La red deberá generalizar usando las zonas de parking donde se pueden aparcar los coches, las texturas de piedra o ladrillo del suelo y paredes, los coches y la puerta negra de metal que da acceso al exterior de la residencia.
- *Ejemplos:* Se pueden ver algunos ejemplos de este nodo en la Figura 2.8.

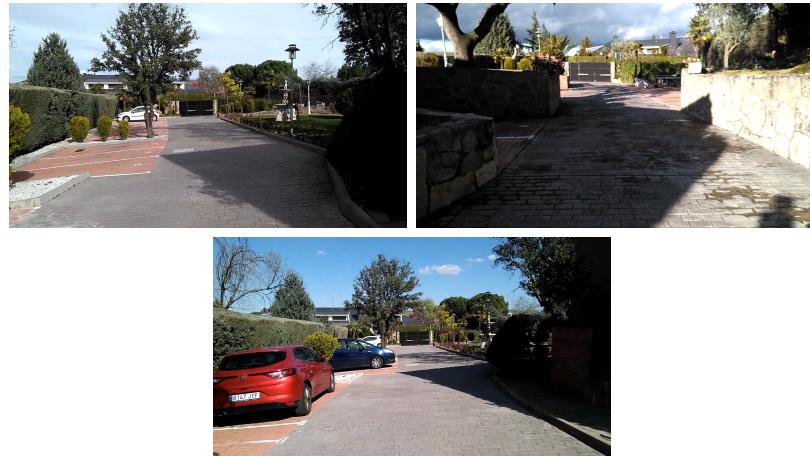


Figura 2.8: Ejemplos de imágenes pertenecientes al nodo 5, el parking

■ Nodo 6 - Biblioteca

- *Descripción:* En el exterior se encuentra la fachada de la biblioteca o zona de estudios de la residencia.
- *Conexiones:* Desde este nodo se puede ir a: el nodo 5, el parking; y el nodo 7, el gimnasio.
- *Landmark:* La fachada de la biblioteca.
- *Generalización:* La red deberá generalizar usando las paredes de piedra que rodean al edificio, las grandes ventanas y puerta de cristal y el cartel que indica que se trata de la biblioteca.
- *Ejemplos:* Se pueden ver algunos ejemplos de este nodo en la Figura 2.9

■ Nodo 7 - Gimnasio



Figura 2.9: Ejemplos de imágenes pertenecientes al nodo 6, la biblioteca

- *Descripción:* En el exterior se encuentra la fachada del gimnasio de la residencia
- *Conexiones:* Desde este nodo se puede ir a: el nodo 3, el comedor; el nodo 4, el edificio de las habitaciones; y el nodo 6, la biblioteca.
- *Landmark:* La fachada del gimnasio.
- *Generalización:* La red deberá generalizar usando la fachada del edificio del gimnasio, las texturas y colores del edificio, el cartel que indica que se trata del gimnasio, los alrededores del gimnasio como la zona de ocio o la mesa de piedra.
- *Ejemplos:* Se pueden ver algunos ejemplos de este nodo en la Figura 2.10



Figura 2.10: Ejemplos de imágenes pertenecientes al nodo 7, el gimnasio

2.4. Navegación

En este proyecto de tesis de fin de máster no contamos con ningún robot físico por lo que la navegación no será una de las prioridades que tendremos que tener en mente para esta tesis.

No obstante este proyecto iría enfocado con el objetivo de que al finalizarlo un robot físico pudiese moverse por el mapa topológico diseñado gracias a nuestra red neuronal, y para ello debemos especificar algunos apartados importantes sobre la navegación en el mapa:

- *Robot físico:* En los terrenos de interior un robot sencillo (con, por ejemplo, un par de ruedas y un sistema odométrico) podría moverse sin ninguna dificultad, pero en este proyecto estamos trabajando en entornos mixtos de interior/exterior y eso complica la navegación pues en el exterior el terreno puede ser más complejo que unas losas de piedra o madera (como podemos encontrarnos en interiores), encontrándonos con distintos terrenos como piedra, tierra, hierba o ladrillo con distinta dificultad de movimiento en cada uno y, por no añadir, que el robot también tendría que enfrentarse a la dificultad de subir escaleras.

Una de las soluciones a este problema sería usar robots polimorfos que pudiesen enfrentarse a estos terrenos abruptos.

Otra posible solución sería usar un UAV (“Unmanned Aerial Vehicle”, un Dron). De esta manera podríamos olvidarnos del problema de los terrenos abruptos pues nuestro robot se movería por el aire.

- *Maniobrabilidad:* Los arcos de los nodos indican desde qué nodo se puede ir a cual nodo, pero además (y aunque no se haya añadido en esta tesis) deberían contener información de cómo realizar las maniobras para viajar entre nodos. Estas maniobras deberían ser lo más simples y robustas posibles para el correcto funcionamiento del robot y para una navegación sin colisiones contra objetos del entorno.

También quiero remarcar que añadir en los arcos el conocimiento de las maniobras no es imprescindible; siempre se puede realizar la navegación con algoritmos reactivos, como el algoritmo bug. También se podría, una vez realizado un mapa interno simbólico de cada nodo, usar algoritmos de resolución de rutas como el A* o Grassfire.

En todo caso, y ya que no contamos con robots físicos para el TFM, estas son simplemente algunas ideas importantes sobre la navegación que podrían implementarse si se decidiera ampliar el trabajo.

También tengo que añadir que, a pesar de que no realicemos un trabajo de navegación con robot físicos, el desarrollo del mapa topológico ha estado siempre enfocado en mejorar el entendimiento de un robot sobre el terreno, influyendo la navegabilidad en la selección de los nodos definidos anteriormente en 2.3 y de los landmarks escogidos para representar cada nodo del mapa topológico.

Capítulo 3

Generación del dataset de imágenes de landmarks para el entrenamiento y evaluación de los reconocedores

3.1. Creación del dataset

3.1.1. Toma de imágenes

3.1.2. Preprocesamiento

3.2. Contenido del Dataset

- Training:

- *Nodo 1*: 4761 imágenes
- **Nodo 2**: 5901 imágenes
- **Nodo 3**: 4140 imágenes
- **Nodo 4**: 5391 imágenes
- **Nodo 5**: 9701 imágenes
- **Nodo 6**: 7342 imágenes
- **Nodo 7**: 4857 imágenes

- Validation:

- **Nodo 1**: 529 imágenes
- **Nodo 2**: 656 imágenes
- **Nodo 3**: 460 imágenes
- **Nodo 4**: 600 imágenes
- **Nodo 5**: 1078 imágenes

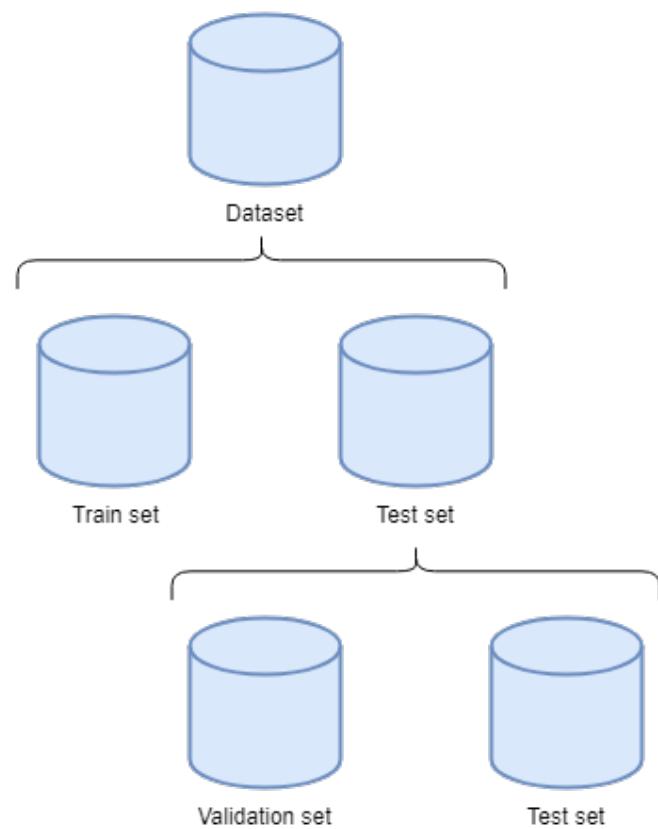


Figura 3.1: Dataset para Redes Neuronales

- **Nodo 6:** 816 imágenes
- **Nodo 7:** 540 imágenes



Figura 3.2: División del dataset para este proyecto

Capítulo 4

Clasificador: Deep Learning + CNNs

4.1. Redes Neuronales

4.2. CNNs - Convolutional Neural Networks

4.3. Técnicas usadas

Capítulo 5

Validation: test estáticos

Capítulo 6

Test: tests dinâmicos

Capítulo 7

Conclusiones y líneas futuras de trabajo

7.1. Conclusiones

7.2. Otras líneas de trabajo

Apéndice A

Anexos

Bibliografía

- [1] Maravall D, de Lope J and Fuentes JP *Vision-based anticipatory controller for the autonomous navigation of an UAV using artificial neural networks*, Neurocomputing, Volume 151, Part 1, 2015, Pages 101-107, ISSN 0925-2312.
- [2] Maravall D, de Lope J and Fuentes JP (2017) *Navigation and Self-Semantic Location of Drones in Indoor Environments by Combining the Visual Bug Algorithm and Entropy-Based Vision*. Front. Neurorobot. 11:46. doi: 10.3389/fnbot.2017.00046