



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
2.3.1. El dilema de concreción versus generalidad	7
2.3.2. Landmarks en detalle	8
2.4. Navegación	13
3. Generación del dataset de imágenes de landmarks para el entrenamiento y evaluación de los reconocedores	15
3.1. Definición del dataset	15
3.2. Toma de imágenes	16
3.3. Tratamiento de las imágenes	16
3.4. Contenido del Dataset	17
3.4.1. Training	17
3.4.2. Validation	18
4. Clasificador: Deep Learning + CNNs	19
4.1. Deep Learning	19
4.1.1. Redes neuronales	19
4.1.2. CNNs - Convolutional Neural Networks	19
4.2. Entrenamiento	19
4.2.1. Herramientas	19
4.2.2. Sistema	19
4.2.3. Arquitectura de la red	19

4.3.	Técnicas usadas	19
4.3.1.	Data augmentation	19
4.3.2.	Dropout	19
4.3.3.	Pooling	20
4.3.4.	Batch normalization	20
5.	Validation: test estáticos	21
5.0.1.	Jupyter notebook	21
5.0.2.	Resultados	21
6.	Test: tests dinámicos	23
6.0.1.	Jupyter notebook	23
6.0.2.	Resultados	23
7.	Conclusiones y líneas futuras de trabajo	25
7.1.	Conclusiones	25
7.2.	Otras lineas 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	8
2.4.	Ejemplos de imágenes pertenecientes al nodo 1, el pasillo de las habitaciones	9
2.5.	Ejemplos de imágenes pertenecientes al nodo 2, las escaleras del hall .	10
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	11
2.8.	Ejemplos de imágenes pertenecientes al nodo 5, el parking	12
2.9.	Ejemplos de imágenes pertenecientes al nodo 6, la biblioteca	12
2.10.	Ejemplos de imágenes pertenecientes al nodo 7, el gimnasio	13
3.1.	Dataset para Redes Neuronales	15
3.2.	División del dataset para este proyecto	17
4.1.	Ejemplo del uso del dropout	20

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.

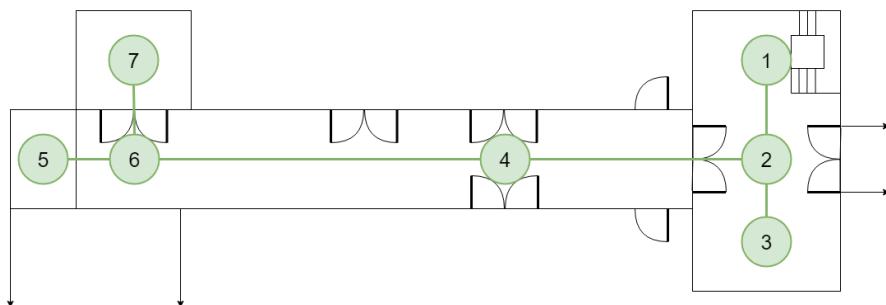


Figura 1.1: Mapa estudiado en la asignatura de robots autónomos

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 2.1.



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 2.3.2 explicaré cada nodo en detalle.

2.3. Landmarks seleccionados

2.3.1. El dilema de concreción versus generalidad

La selección de los landmarks y sus respectivas imágenes es una etapa crucial en la generación manual de un mapa topológico visual basado en grafos visuales ya que plantea un dilema crítico entre la concreción o especialización por una parte y la generalidad por otra de las imágenes asociadas a los landmarks del mapa.

Se puede afirmar que los landmarks “generalistas” tienen la ventaja de una mayor simplicidad y ergonomía antropomórfica de los correspondientes mapas topológicos visuales (en el sentido de ser más intuitivos o perceptualmente comprensibles para un usuario humano, lo cual facilita el desarrollo del diseño y testeo de los mapas y de las pruebas del navegador, así como la posible interacción humano-robot en un entorno de navegación física del robot con intervención o cooperación humana).

No obstante, frente a la citada ventaja “antropomórfica” de los landmarks “generalistas”, éstos también plantean un reto mayor en el desarrollo del módulo de reconocimiento ya que las imágenes asociadas a los landmarks “generalistas” son más complejas que las imágenes asociadas a landmarks “particularistas”. Se puede afirmar que los landmarks generalistas requieren reconocedores de elevada capacidad de generalización, contrariamente a los reconocedores de landmarks particularistas, mucho más especializados.

En [1] puede observarse que se han definido los landmarks mediante la selección de objetos concretos y bien diferenciados. Este suele ser el procedimiento más empleado: por ejemplo, en [2] se seleccionaron una puerta o un extintor. En el presente trabajo no se ha optado por este habitual enfoque particularista de selección de los landmarks centrados en objetos concretos y bien diferenciados, sino que hemos optado por un enfoque más generalistas.

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.

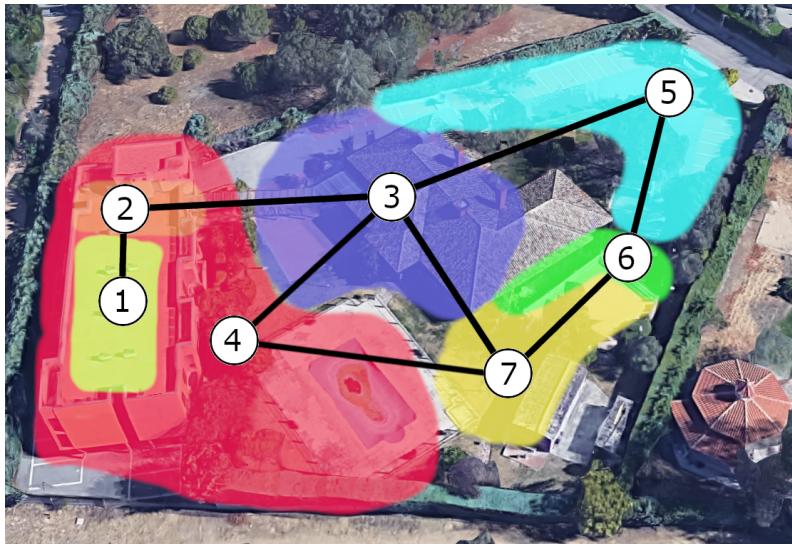


Figura 2.3: Áreas de estudio de cada nodo

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 2.3.2 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.

En resumen, en este trabajo hemos decidido investigar la problemática asociada al uso de landmarks generalistas y por ese motivo nos hemos decantado hacia la experimentación con los métodos basados en redes neuronales convolucionales y aprendizaje profundo, en principio con mayor capacidad de generalización y reconocimiento complejo que otras técnicas de naturaleza más especializada.

2.3.2. 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

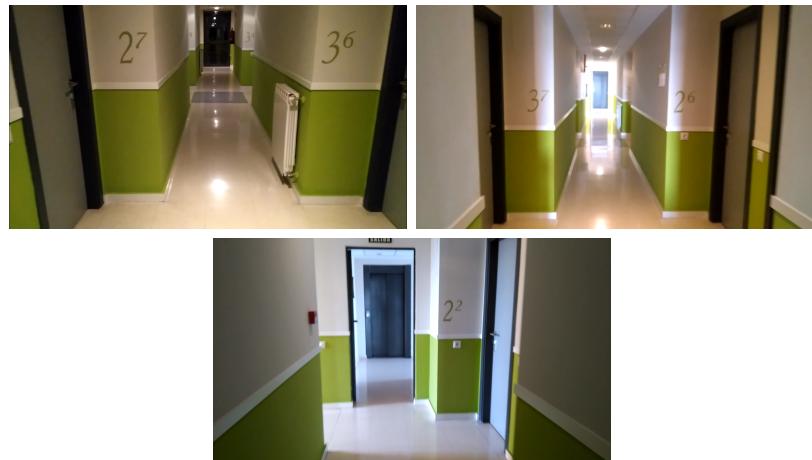


Figura 2.4: Ejemplos de imágenes pertenecientes al nodo 1, el pasillo de las habitaciones

■ 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

■ 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.

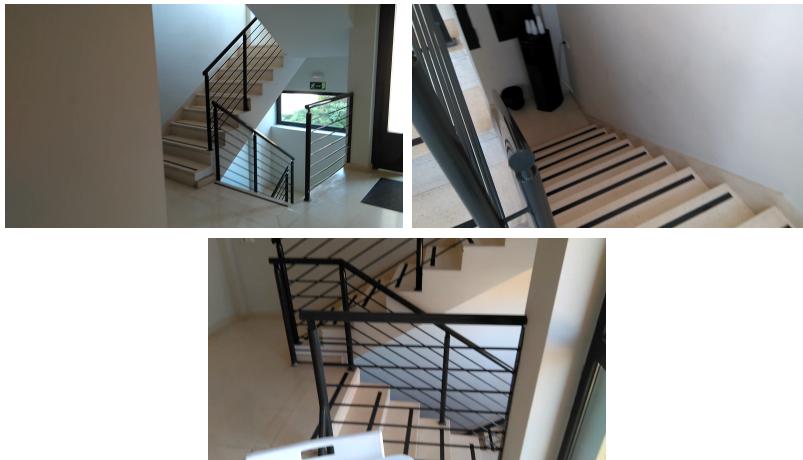


Figura 2.5: Ejemplos de imágenes pertenecientes al nodo 2, las escaleras del hall

- *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



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

■ Nodo 4 - Edificio de las habitaciones

- *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.

■ 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.

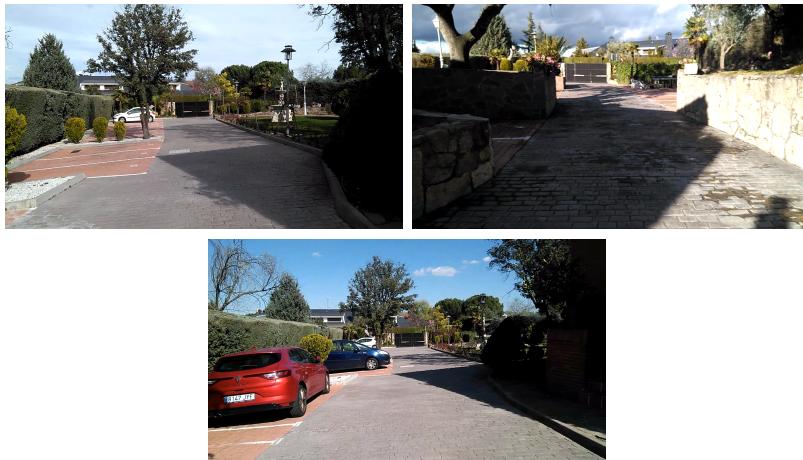


Figura 2.8: Ejemplos de imágenes pertenecientes al nodo 5, el parking

- *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



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

■ Nodo 7 - Gimnasio

- *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 que nodo se puede ir a cual nodo, pero además (y aunque no se haya añadido en esta tesis) deberían contener información de como 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 2.3.2 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. Definición del dataset

El primer paso para empezar a entrenar la red neuronal era diseñar la organización estructural para guardar nuestro dataset.

De esta manera, dividí el dataset en “train”, “validation” y “test” como puede verse en la Figura 3.1.

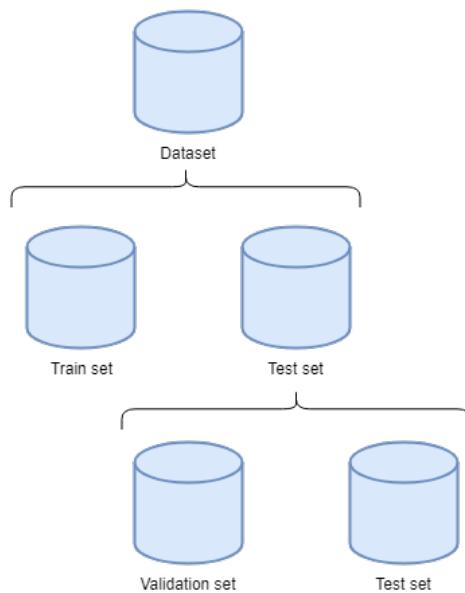


Figura 3.1: Dataset para Redes Neuronales

Esta es la organización habitual para entrenar una red neuronal. Representé esta organización en una jerarquía de carpetas de la siguiente manera:

- *data*: el dataset.
 - *train*: las imágenes de entrenamiento de la red neuronal.
 - *Nodo1*: las imágenes de entrenamiento del nodo 1.
 - ...
 - *Nodo7*: las imágenes de entrenamiento del nodo 2.
 - *validation*: las imágenes de validación de la red neuronal.
 - *Nodo1*: las imágenes de validación del nodo 1.
 - ...
 - *Nodo7*: las imágenes de validación del nodo 2.
- *test*: los videos de test finales.
 - *originals*: los videos originales de test sobre los que se les pasará frame a frame la red neuronal para comprobar su rendimiento.
 - *results*: los videos generados por los scripts en los que se clasifica frame a frame un video de la carpeta “*originals*”.

3.2. Toma de imágenes

El proceso de toma de imágenes es muy simple; he grabado una serie de vídeos y he guardado cada uno de los frames de esos vídeos como una imagen.

Para realizar esto me he apoyado en un cuaderno jupyter en python que he creado llamado “*Dataset Refinement.ipynb*”.

Para usar el cuaderno jupyter lo primero que hay que hacer es renombrar los vídeos de los que queremos extraer imágenes. Para ello debemos llamarlos igual que el nodo al que pertenece el vídeo junto a una barra baja y a un número para poner un orden en el que se irá accediendo a estos vídeos. De esta manera tendremos vídeos con nombre “Nodo{Número del nodo} _{Otro número para indicar el orden de acceso}”.

El cuaderno jupyter va leyendo vídeo a vídeo los frames y los va guardando en la carpeta correspondiente a su nodo, no sin antes realizar las operaciones de conversión que serán explicadas en el siguiente apartado 3.3.

3.3. Tratamiento de las imágenes

La cámara con la que se realizaron los vídeos generó imágenes de tamaño 1920x1080. No podemos pasarle a la red neuronal imágenes tan grandes para que entrene con ellas pues el tiempo que tardaría en entrenar podría ser de meses.

He reducido el tamaño de las imágenes y he probado con distintos formatos, siempre manteniendo la imagen con un tamaño rectangular tal y como se tomaron con la cámara y siempre intentando mantener los objetos de la imagen reconocibles.

Se probaron distintos formatos: 960 x 540, 480 x 270 y 240 x 135, pero seguía siendo demasiado grande y la red tardaba aún bastante tiempo en realizar el entrenamiento, por lo que se optó por disminuir el tamaño aún más hasta los 128 x 72, en los que la red solo tardaba una hora en realizar el entrenamiento con el dataset mientras se seguían manteniendo los bordes que delimitaban los objetos aún visibles.

3.4. Contenido del Dataset

El primer dataset con el que empecé a entrenar la red neuronal contaba con 500 imágenes para entrenamiento y 50 para validación por nodo y, a medida que la red neuronal iba mostrándome sus resultados, había que añadir al dataset más imágenes de algunos escenarios que no era capaz de generalizar.

El dataset final de este proyecto está compuesto por 46.772 imágenes de las que 42.093 han sido seleccionadas para entrenamiento y 4.679 para validación. Esto supone que el 90 % de las imágenes totales del dataset han sido usadas para entrenamiento y el 10 % de las imágenes totales del dataset han sido usadas para la validación. Dicha división ha sido representada en la Figura 3.2.

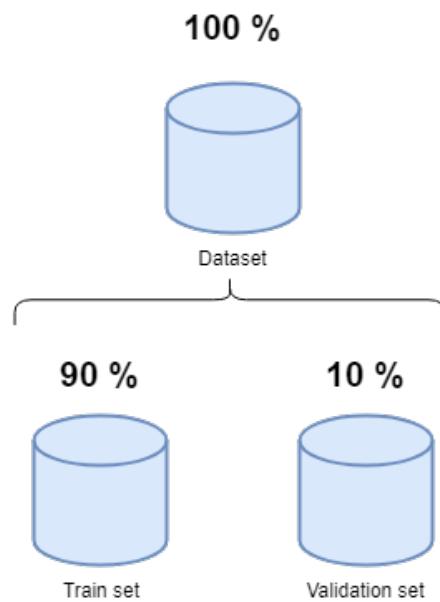


Figura 3.2: División del dataset para este proyecto

3.4.1. Training

Ya que tenemos alrededor de 42.000 imágenes y 7 nodos, para el entrenamiento idílicamente a cada nodo le tocárían, por media, unas 6.000 imágenes. No obstante, realizando iteraciones para mejorar los resultados de la red neuronal me he visto

obligado a tomar más fotos de algunos nodos que no generalizaban bien. En cambio, otros nodos generalizaban bien desde un principio y no necesitaban tantas imágenes para llegar a buenos resultados.

El resultado final ha sido el siguiente:

- *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

Los nodos han necesitado un mínimo de 4.000 imágenes para generalizar y converger en buenos resultados mientras que el nodo 5 ha necesitado muchas más imágenes para poder generalizar. En el Capítulo 4 hablaremos en más detalle del proceso de entrenamiento y de las dificultades en la generalización.

3.4.2. Validation

En lo que respecta a la validación, simplemente he tomado un 10 % de imágenes aleatorias de cada nodo del dataset final. El resultado es el siguiente:

- *Nodo 1*: 529 imágenes
- *Nodo 2*: 656 imágenes
- *Nodo 3*: 460 imágenes
- *Nodo 4*: 600 imágenes
- *Nodo 5*: 1078 imágenes
- *Nodo 6*: 816 imágenes
- *Nodo 7*: 540 imágenes

Capítulo 4

Clasificador: Deep Learning + CNNs

4.1. Deep Learning

4.1.1. Redes neuronales

4.1.2. CNNs - Convolutional Neural Networks

4.2. Entrenamiento

4.2.1. Herramientas

4.2.2. Sistema

4.2.3. Arquitectura de la red

4.3. Técnicas usadas

A continuación explicaré algunas técnicas usadas en la elaboración de esta red neuronal que son merecedoras de ser explicadas pues intervienen directamente en el proceso de generalización de la red y contribuyen a un incremento en la tasa de aciertos en los test finales.

4.3.1. Data augmentation

Muchas veces los datos con los que contamos para el entrenamiento no son suficientes. Es por eso por lo que

4.3.2. Dropout

Mediante el Dropout lo que estoy haciendo es

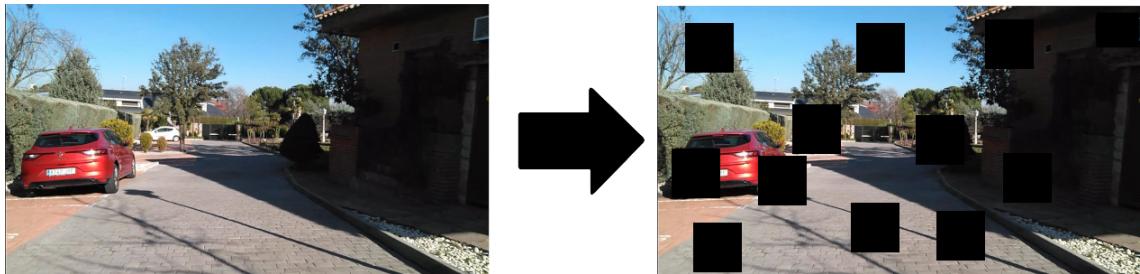


Figura 4.1: Ejemplo del uso del dropout

4.3.3. Pooling

4.3.4. Batch normalization

Capítulo 5

Validation: test estáticos

5.0.1. Jupyter notebook

5.0.2. Resultados

Capítulo 6

Test: tests dinâmicos

6.0.1. Jupyter notebook

6.0.2. Resultados

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
- [3] Krizhevsky, Alex & Sutskever, Ilya & E. Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.
- [4] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14). IEEE Computer Society, Washington, DC, USA, 580-587. DOI: <https://doi.org/10.1109/CVPR.2014.81>