

Proyecto de Grado

Aprendizaje fin a fin para la conducción autónoma de vehículos domésticos usando visión artificial y redes neuronales convolucionales

Postulante: Jose Eduardo Laruta Espejo

Asesor: Javier Sanabria

14 de junio de 2019

Universidad Mayor de San Andrés - Facultad de Ingeniería

Contenido

1. Introducción
2. Fundamento Teórico
3. Subsistema de Control y Actuación
4. Subsistema de Adquisición de Datos y Entrenamiento
5. Subsistema de Inferencia y control autónomo
6. Análisis de Resultados
7. Conclusiones y Recomendaciones

Introducción

Planteamiento del problema

Un sistema de conducción autónomo cuenta con la siguiente arquitectura:

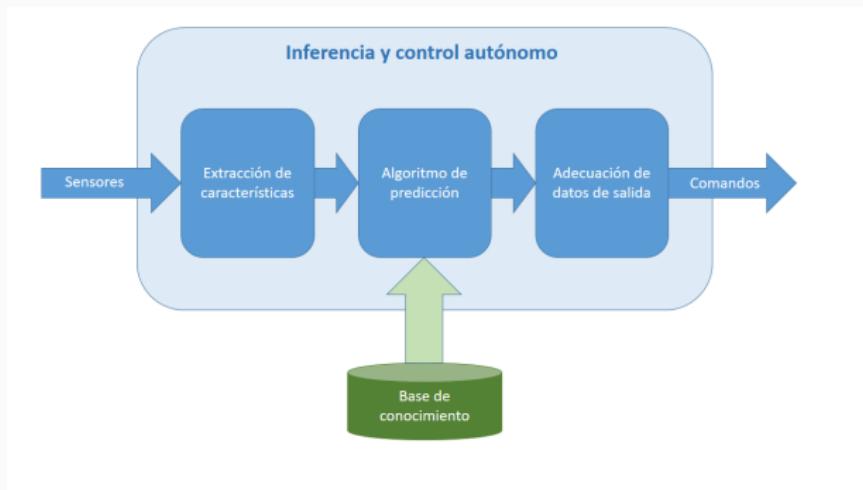


Figura 1: Componentes del subsistema de inferencia y control autónomo tradicional. Fuente: Elaboración propia.

Planteamiento del problema

El sistema presenta las siguientes dificultades:

- Conocimiento experto en cada etapa.
- Poca flexibilidad en el diseño y la implementación.
- Tiempo y recursos necesarios elevados.



Objetivo General

Diseñar un sistema de aprendizaje “*fin a fin*” capaz de generar de comandos de control de vehículos domésticos basado en visión artificial y **redes neuronales convolucionales** para facilitar el diseño e implementación de sistemas de conducción autónoma.

Objetivos Específicos

- Estudiar los aspectos concernientes a los sistemas de conducción autónoma y sistemas de aprendizaje.
- Analizar los requerimientos.
- Diseñar la arquitectura del sistema.
- Diseñar el subsistema de control y actuación.
- Diseñar el subsistema de adquisición de datos y entrenamiento.
- Diseñar el subsistema de inferencia y control autónomo.
- Analizar los resultados del entrenamiento e implementación.
- Realizar pruebas de rendimiento y análisis comparativos.

Fundamento Teórico

Sistemas de Conducción Autónoma

Un sistema de conducción autónoma tiene la capacidad de operar un vehículo de forma completamente autónoma.



Figura 2: Vehículos del *Grand Challenge* de 2005: (a) Stanley (1er lugar), (b) Sandstorm (2do lugar). Vehículos del *Urban Challenge* de 2007: (c) Boss (1er lugar) (d) Junior (2do lugar). Fuente: [1]

Niveles de Autonomía SAE

Se han definido distintos niveles de autonomía por parte de la Sociedad de Ingenieros de Automoción (SAE).

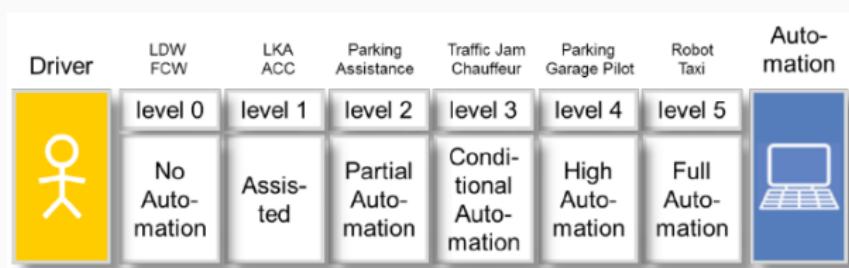


Figura 3: Niveles de automatización en la conducción según SAE. Fuente: researchgate

Visión Artificial

Es un campo de las ciencias de la computación que intenta reproducir las capacidades de una persona para entender imágenes.

Aplicaciones:

- Reconocimiento óptico de caracteres (OCR).
- Inspección automática.
- Ventas.
- Reconstrucción de modelos en 3D (fotogrametría).
- Imagenología médica.
- Seguridad automotriz.
- Seguridad y vigilancia.

Visión Artificial

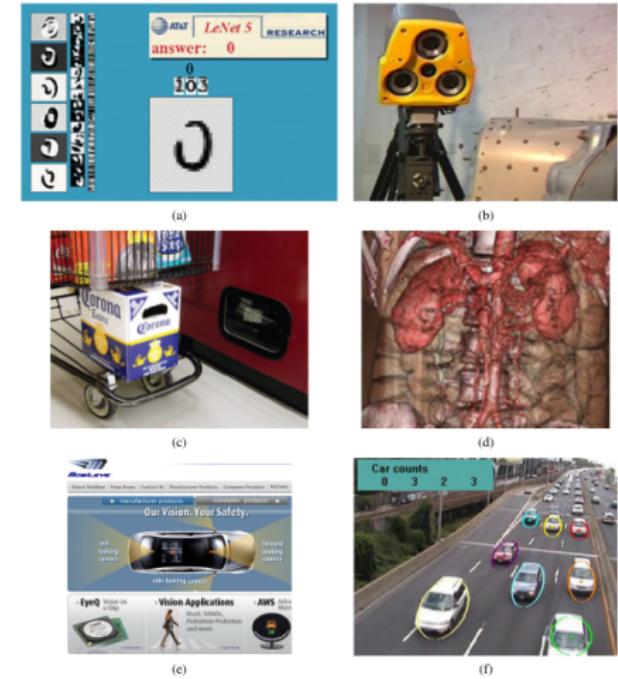


Figura 4: Algunas aplicaciones de la visión artificial. Fuente: [2]

Redes Neuronales Artificiales

Las redes neuronales artificiales son un modelo matemático de cómputo muy útiles en la aproximación de funciones y reconocimiento de patrones.

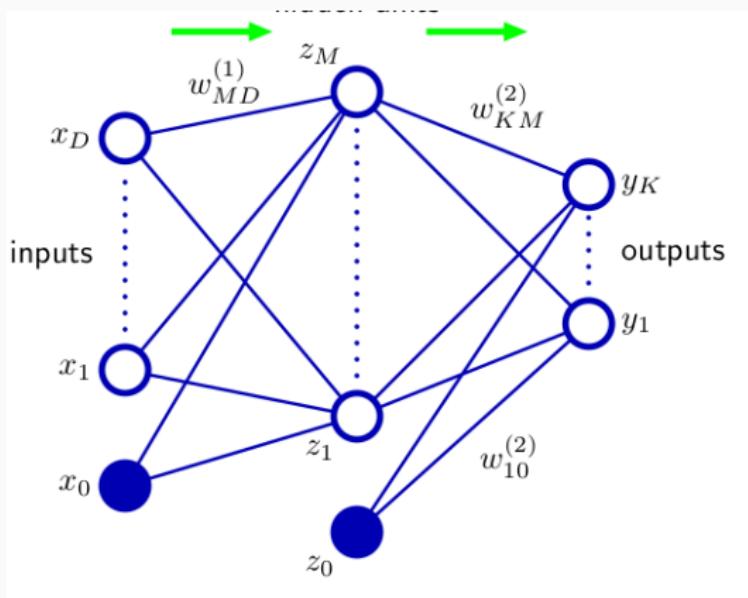
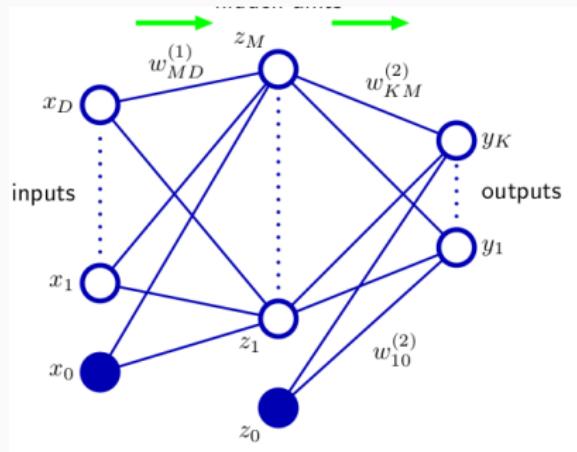


Figura 5: Diagrama de la red neuronal de dos capas. Fuente: [3]

Redes Neuronales Artificiales



Ecuación correspondiente:

$$y_k(\mathbf{x}, \mathbf{W}) = \sigma \left(\sum_{j=1}^M w_{kj}^{(2)} h \left(\sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right) \quad (1)$$

Función de Costo

La función de costo permite definir el rendimiento de la red con respecto a las predicciones esperadas a la salida.

Dado un conjunto de entrenamiento compuesto por un conjunto de vectores de entrada $\{\mathbf{x}_n\}$, donde $n = 1, \dots, N$, junto con un conjunto de vectores objetivo $\{\mathbf{t}_n\}$ el objetivo es **minimizar la función**:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(\mathbf{x}_n, \mathbf{w}) - t_n\}^2 \quad (2)$$

Descenso de Gradiente

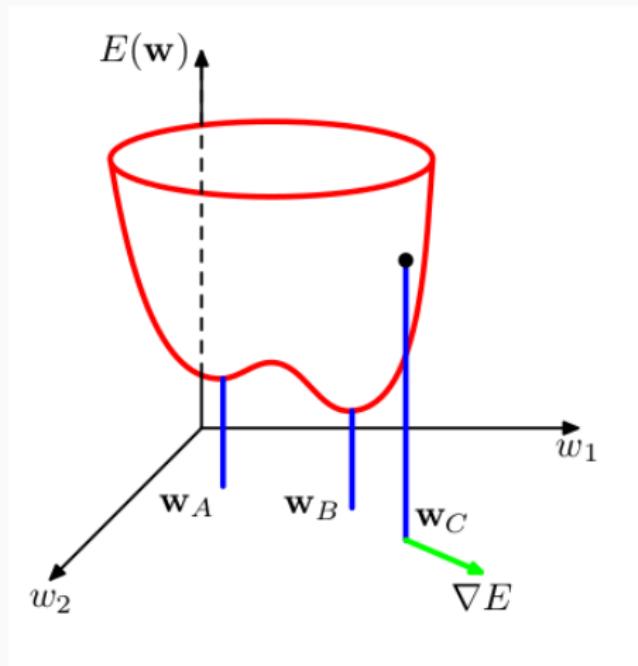


Figura 6: Visualización de la función de error $E(\mathbf{w})$ como una superficie. El punto \mathbf{w}_A es un mínimo local y el punto \mathbf{w}_B es el mínimo global. Fuente: [3]

Redes Neuronales Convolucionales

Las redes convolucionales usan la función de convolución para procesar la información en cada capa oculta.

La operación de convolución para señales digitales se define de la siguiente manera:

$$s(t) = (x * w)(t) = \sum_{\tau=-\infty}^{\infty} x(\tau)w(t - \tau) \quad (3)$$

Para el caso de una imagen en 2D, la convolución se define como:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad (4)$$

Redes Neuronales Convolucionales

La operación de convolución se puede aplicar a imágenes (matrices multidimensionales):

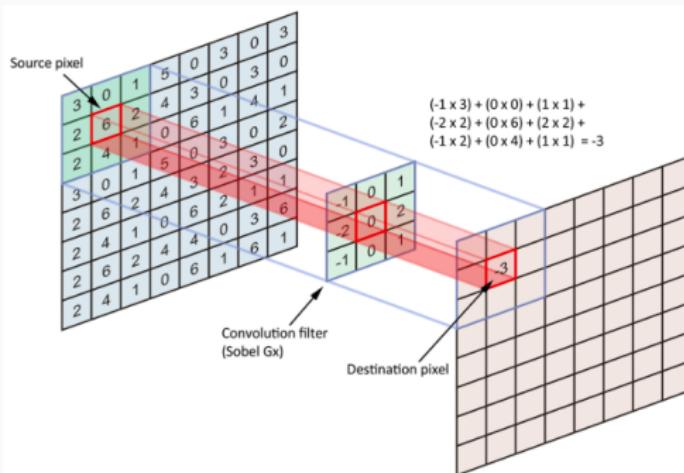


Figura 7: Visualización de la operación de convolución sobre una imagen digital. Fuente: [4]

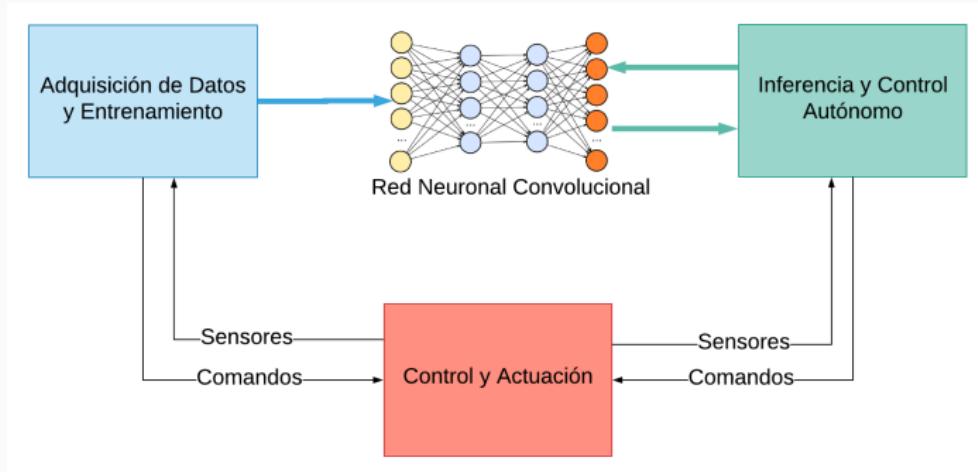
Sistemas de Aprendizaje Fin a Fin

Los sistemas de aprendizaje fin a fin son modelos de aprendizaje que tratan ciertos problemas **de manera conjunta** en lugar de separarlos en módulos o etapas de procesamiento.

Las representaciones internas de un sistema fin a fin se encargan de abstraer el conocimiento en base a la experiencia previa presentada en el conjunto de datos y no necesita de un experto que atienda el desarrollo de las mismas.

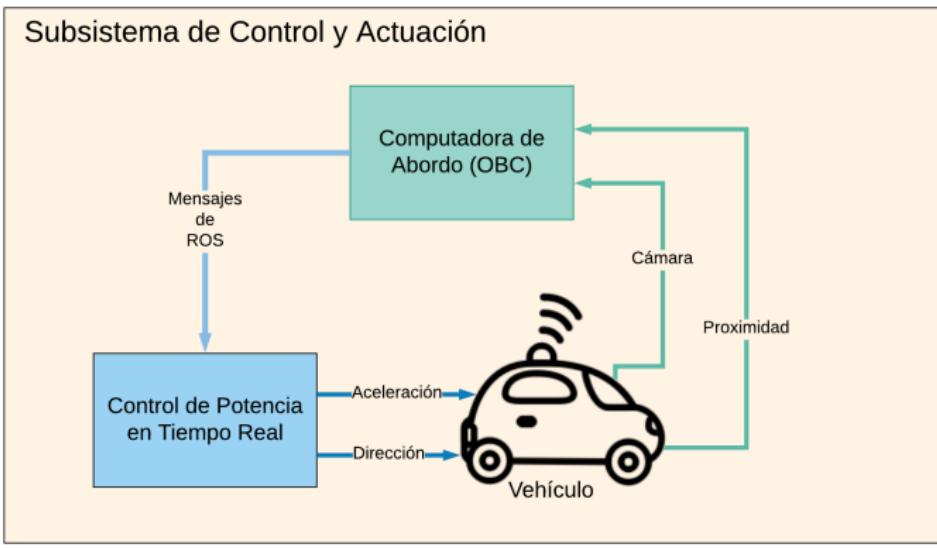
Ingeniería del proyecto

Arquitectura del Sistema



Subsistema de Control y Actuación

Esquema

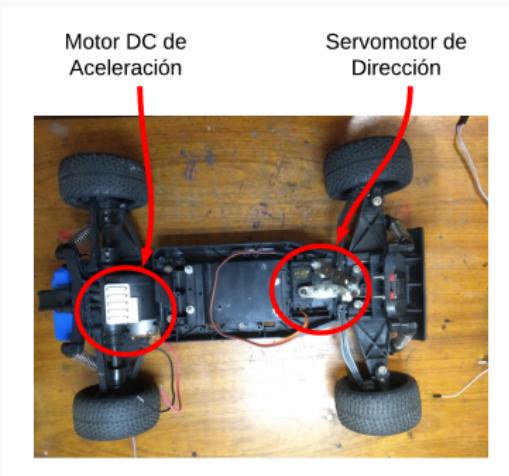


Este subsistema cuenta con módulos que tienen tres principales responsabilidades.

- Ejecutar un control de tiempo real en los actuadores disponibles para la locomoción del vehículo a través de un sistema embebido.
- Brindar una plataforma para la adquisición de los datos de los sensores (la cámara y el sensor de proximidad).
- Brindar una plataforma de desarrollo para el control autónomo del vehículo mediante ROS.

Módulo de potencia en tiempo real

Se encarga del control de los actuadores mediante un sistema embebido que tiene capacidades de tiempo real.



Módulo de potencia en tiempo real

Se encarga del control de los actuadores mediante un sistema embebido que tiene capacidades de tiempo real.

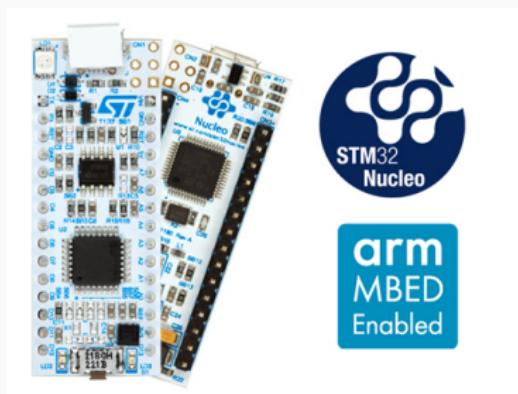


Figura 8: Placa de desarrollo Nucleof303k8 de ST Microelectronics. Fuente: [5]

Módulo de la computadora de abordo (OBC)

La OBC se encarga del control de alto nivel del vehículo. Sus principales tareas son:

- Recepción de comandos de control remoto.
- Comunicación de los nodos de procesamiento del sistema.
- Envío de comandos de control al módulo de tiempo real.

Módulo de la computadora de abordo (OBC)

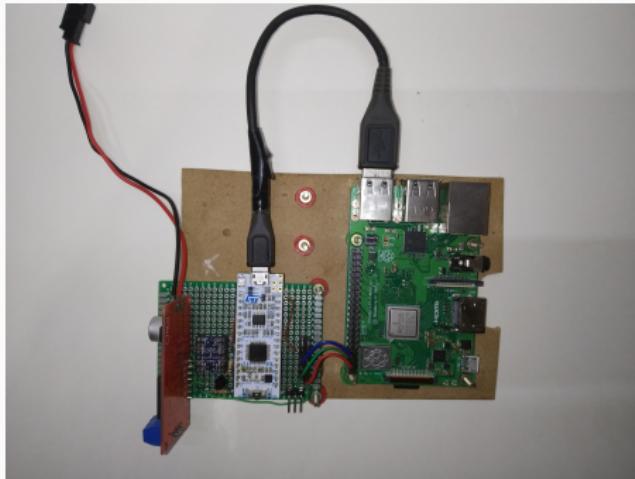


Figura 9: Fotografía de la OBC. Fuente: Elaboración propia.

Módulo de la computadora de abordo (OBC)

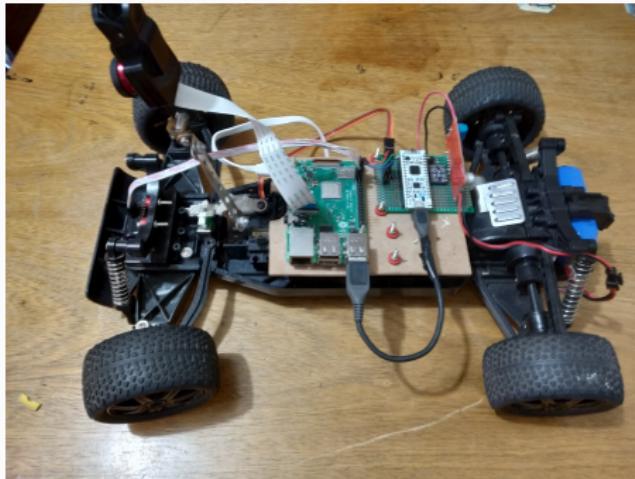


Figura 10: Fotografía del prototipo. Fuente: Elaboración propia.

Módulo de la computadora de abordo (OBC)

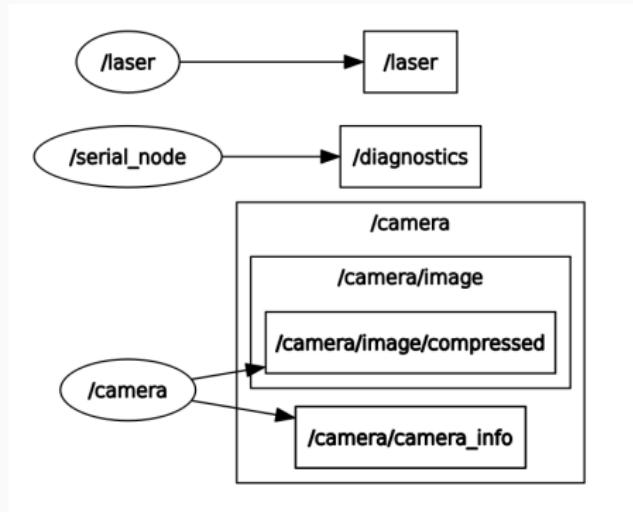
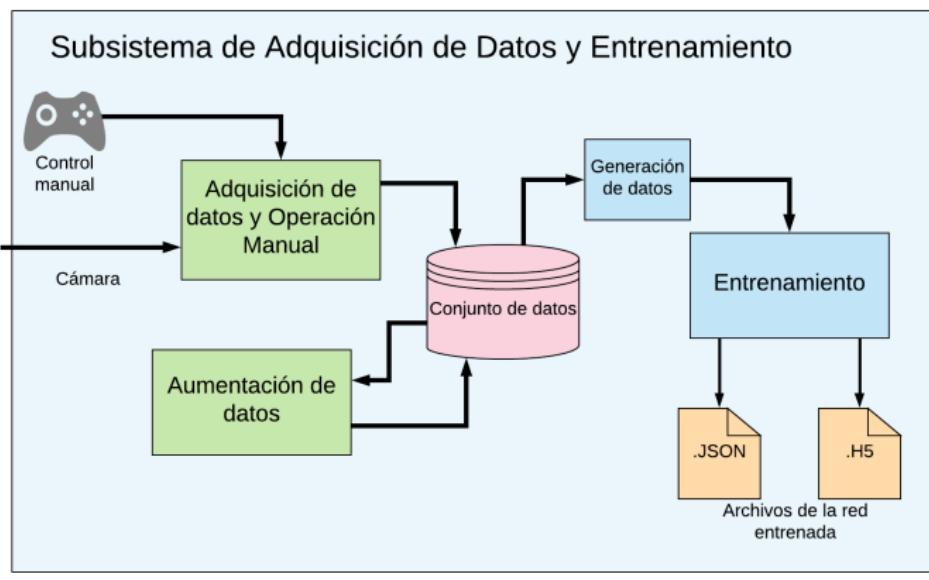


Figura 11: Interacción de los nodos en la OBC. Fuente: Elaboración propia.

Subsistema de Adquisición de Datos y Entrenamiento

Esquema



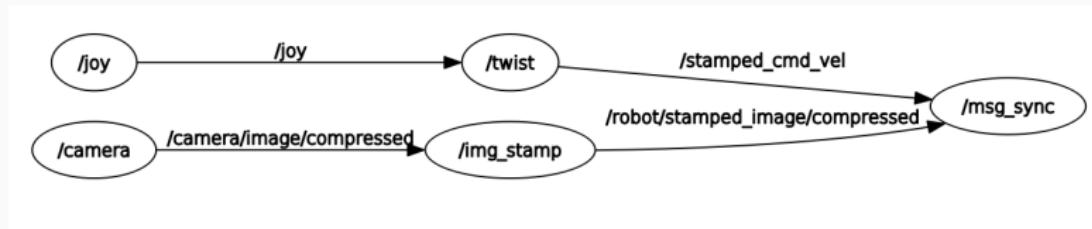
Tareas

- Generar un conjunto de datos o *dataset* para el entrenamiento y validación de la red neuronal.
- Entrenar una red neuronal a partir del conjunto de datos y ciertos parámetros previamente definidos.

Módulo de Adquisición de Datos y Operación Manual

Se realiza la adquisición de datos mientras se opera el vehículo manualmente.

Este modo de funcionamiento permite almacenar datos para el entrenamiento.



Módulo de Adquisición de Datos y Operación Manual



Figura 12: Imágenes capturadas por la cámara en una sesión de entrenamiento. Fuente: Elaboración propia.

Módulo de Aumentación de Datos

Una de las desventajas de las redes neuronales es que se necesita una cantidad sustancialmente superior de datos que en otros algoritmos de visión artificial.

Por su parte, la obtención del conjunto de datos es uno de los procesos más costosos en tiempo y recursos en un sistema de aprendizaje.

Módulo de Aumentación de Datos

Transformación	Descripción
Rotación	Rotación de la imagen en un ángulo aleatorio entre -20 y 20 grados.
Desplazamiento Vertical	Desplazamiento de la imagen de forma vertical un valor aleatorio entre 0 y 20 %.
Espejo Horizontal	Espejado de la imagen horizontalmente de forma aleatoria.

Tabla 1: Transformaciones realizadas en la aumentación de datos. Fuente:
Elaboración propia.

Módulo de Aumentación de Datos

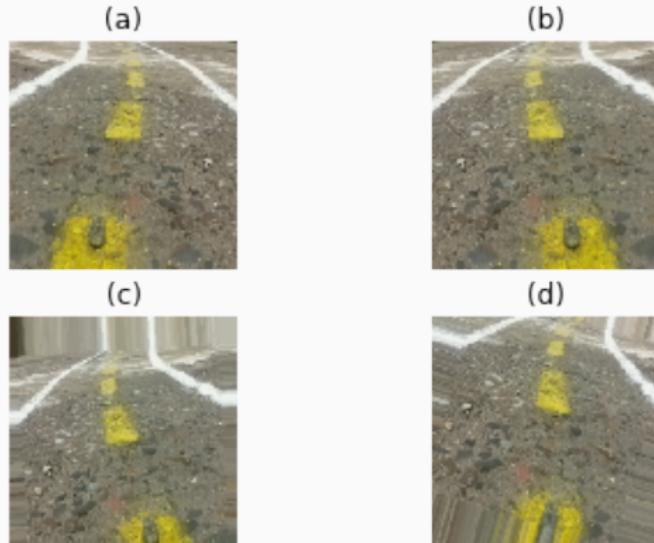


Figura 13: Ejemplos del proceso de aumentación de datos: (a) imagen original, imagen espejada horizontalmente, (c) imagen desplazada verticalmente, (d) imagen rotada. Fuente: Elaboración propia.

Módulo de Generación de Datos

Este módulo genera conjuntos pequeños de imágenes que se cargan a la memoria de manera progresiva para el entrenamiento debido a que no se puede cargar todo el conjunto de entrenamiento de una sola vez.

Módulo de Entrenamiento

Este módulo realiza las siguientes tareas:

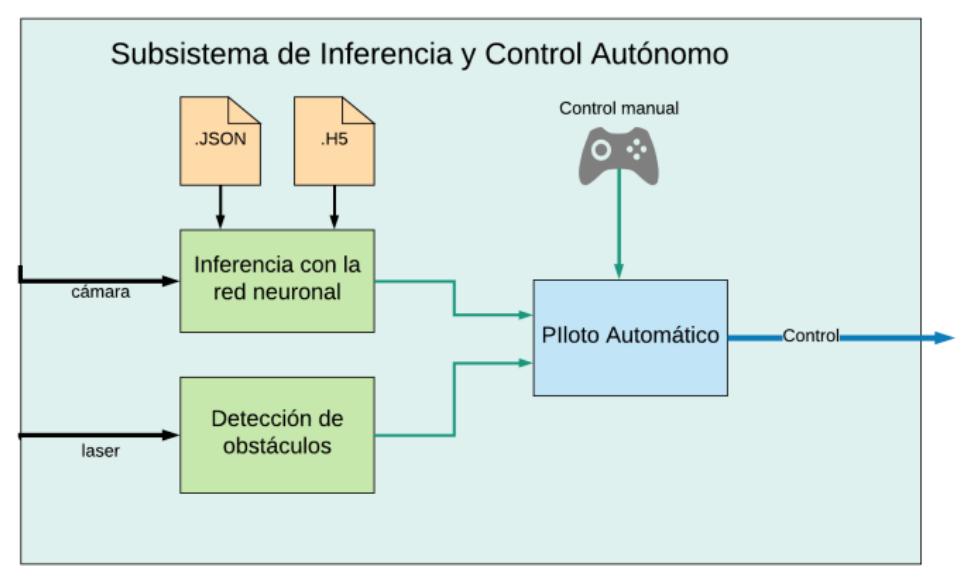
- Cargar el dataset y el modelo de la red.
- Definir hiperparámetros.
- Ejecutar y monitorear el entrenamiento.
- Salvaguardar parámetros.
- Generar reportes del entrenamiento.

El módulo tiene como salida dos archivos:

- **Arquitectura de la red neuronal:** Se almacena la información acerca de la arquitectura de la red en un archivo con formato *JSON* donde se puede encontrar la información acerca de las dimensiones de las capas ocultas, cantidad de unidades por capa y dimensiones de cada unidad en la red.
- **Pesos entrenados de la red neuronal:** Se almacena también los valores de todos los parámetros o pesos de la red neuronal, resultado del entrenamiento. Estos valores están almacenados en un archivo con extensión *H5*.

Subsistema de Inferencia y control autónomo

Esquema

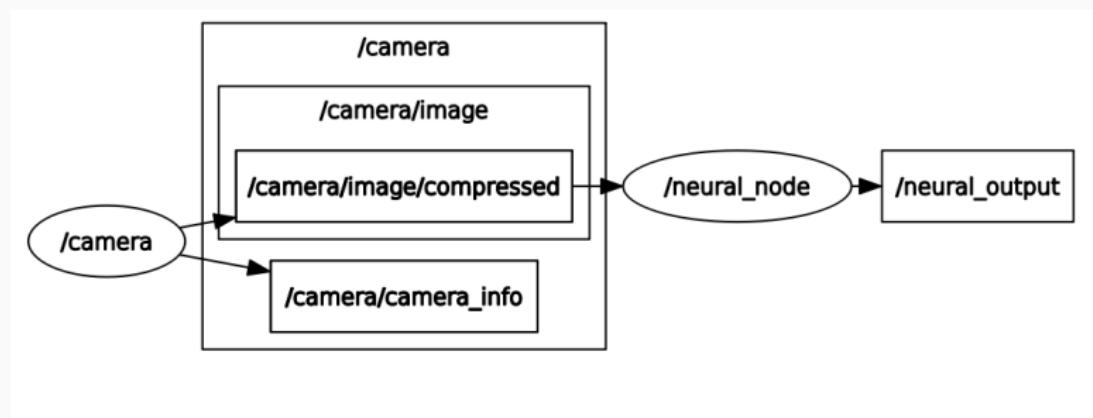


El subsistema debe cumplir las siguientes tareas:

- Cargar y ejecutar el modelo de predicción implementado en la red neuronal entrenada en el subsistema de adquisición de datos y entrenamiento para la generación de comandos de dirección del vehículo.
- Ejecutar un algoritmo de control para la aceleración basado en la detección de obstáculos presentes frente al vehículo.
- Ejecutar un algoritmo de arbitraje que combine ambos sistemas de control en conjunto con un control manual de respaldo que será referido como el *piloto automático*.

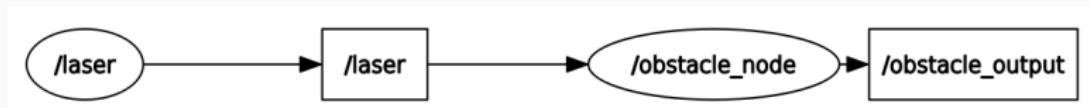
Módulo de Inferencia Neuronal

Este módulo tiene la tarea de ejecutar la tarea de predicción del comando de control de dirección con la red neuronal entrenada por el módulo de entrenamiento del subsistema de adquisición de datos y entrenamiento.



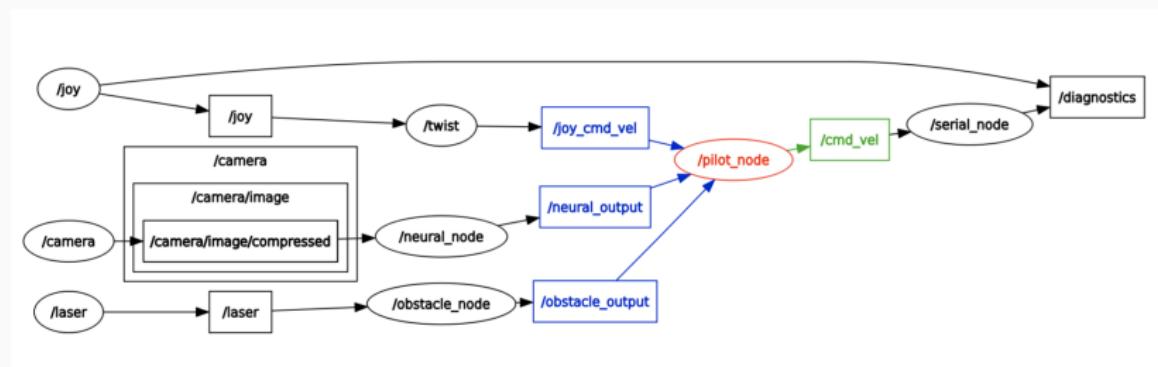
Módulo de detección de obstáculos

El módulo de detección de obstáculos tiene la finalidad de controlar la aceleración del vehículo basado en la presencia de obstáculos físicos frente al mismo. Se basa en las mediciones de proximidad realizadas por el sensor de proximidad montado en frente del vehículo.



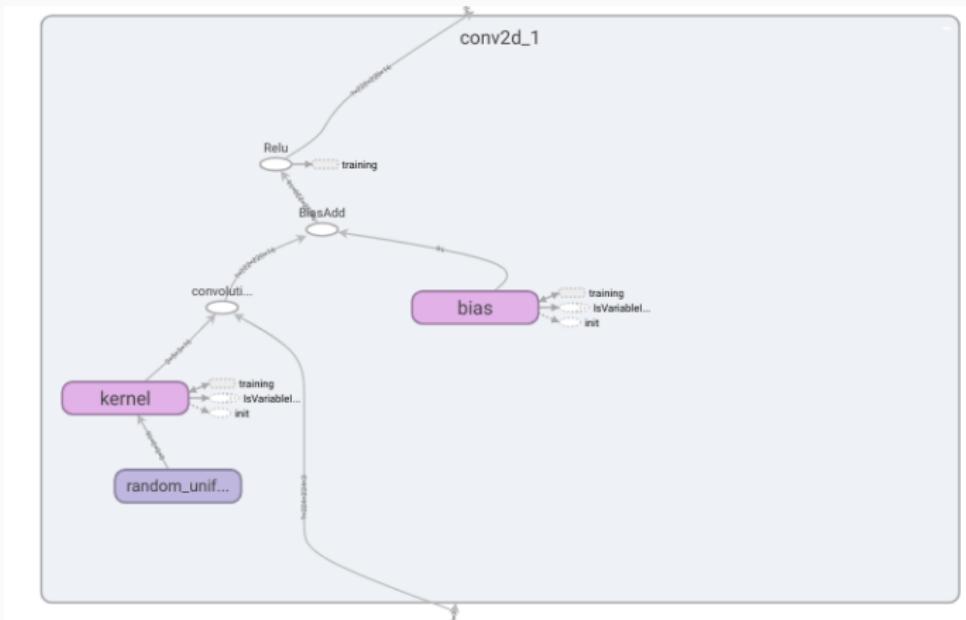
Módulo del piloto automático

Este módulo se encarga de arbitrar los comandos de control provenientes de las distintas fuentes utilizadas en el sistema. Se suscribe a cada tópico correspondiente y redirecciona los mensajes para el control del vehículo mediante mensajes.

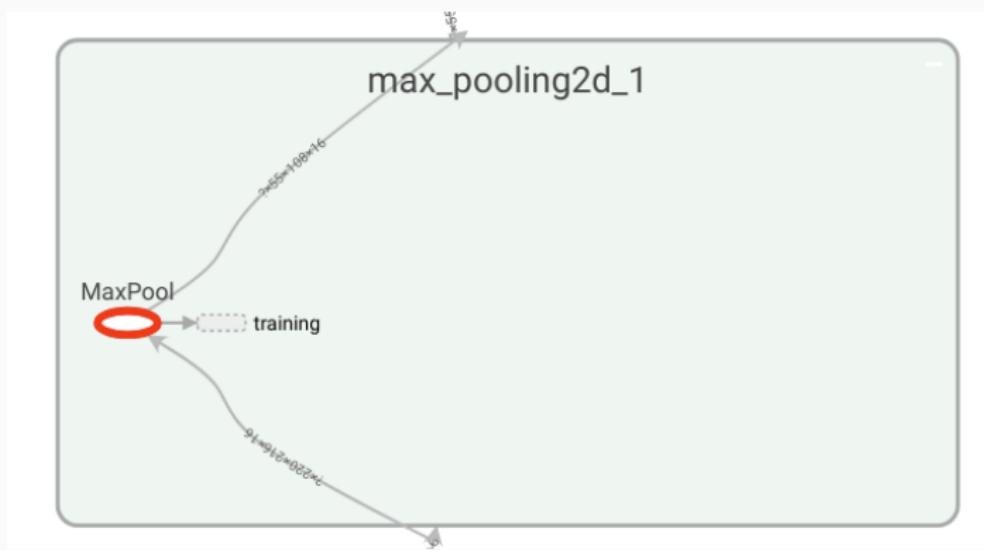


Arquitectura de la Red Neuronal

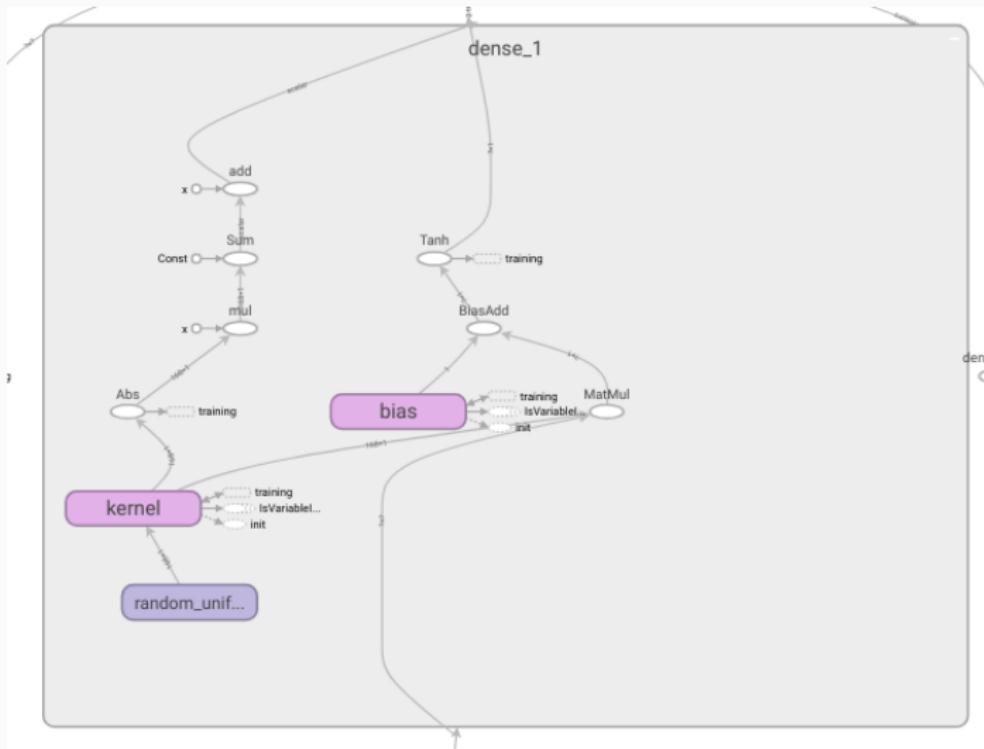
Capas Convolucionales



Pooling



Capas Densamente Conectadas



Funciones de Activación - RELU

$$g(x) = \max\{0, x\} \quad (5)$$

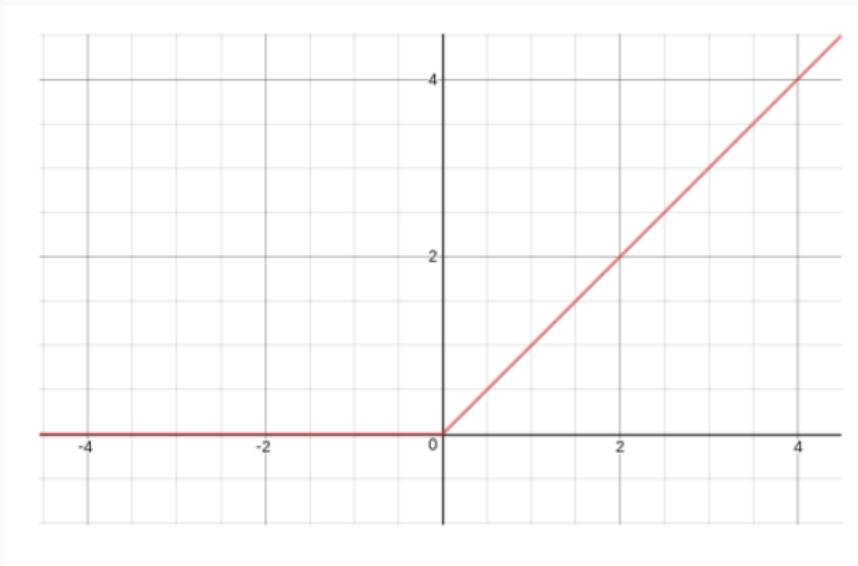


Figura 14: Gráfico de la función ReLU. Fuente: [6]

Funciones de Activación - Tanh

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (6)$$

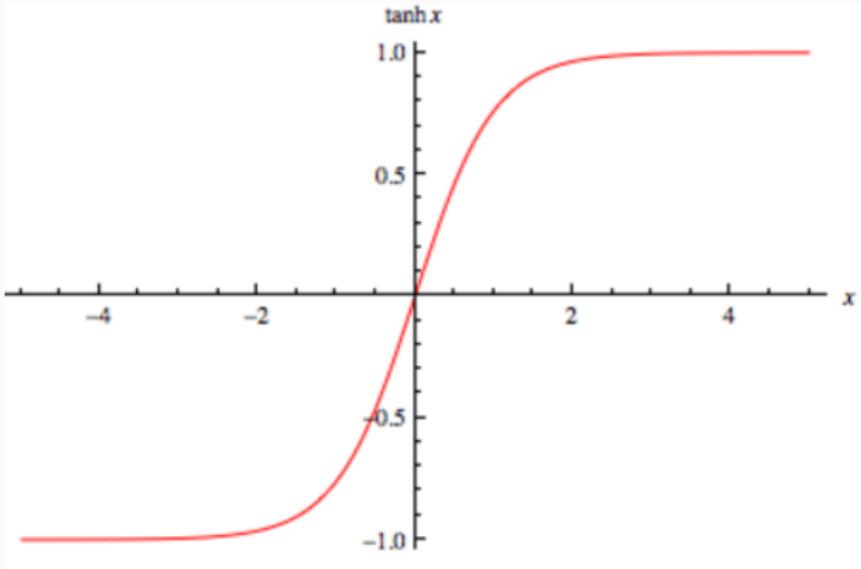


Figura 15: Gráfico de la función tangente hiperbólica $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.
Fuente: [6]

Función de Costo

Se usa la función de **error cuadrático medio** como función de costo para la tarea de regresión, definida por:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(\mathbf{x}_n, \mathbf{w}) - t_n\}^2 \quad (7)$$

Algoritmo de Optimización

Se ha utilizado el algoritmo **ADAM** [7] basado en momentos de primer y segundo orden.

$$\begin{aligned} m_w^{(\tau+1)} &= \beta_1 m_w^{(\tau)} + (1 - \beta_1) \nabla_w E^{(\tau)} \\ v_w^{(\tau+1)} &= \beta_2 m_w^{(\tau)} + (1 - \beta_2) (\nabla_w E^{(\tau)})^2 \\ \hat{m}_w &= \frac{m_w^{(\tau+1)}}{1 - (\beta_1)^{(\tau+1)}} \\ \hat{v}_w &= \frac{v_w^{(\tau+1)}}{1 - (\beta_2)^{(\tau+1)}} \\ w^{(\tau+1)} &= w^{(\tau)} - \alpha \frac{\hat{m}_w}{\sqrt{\hat{v}_w} + \epsilon} \end{aligned} \tag{8}$$

Implementación de Arquitecturas

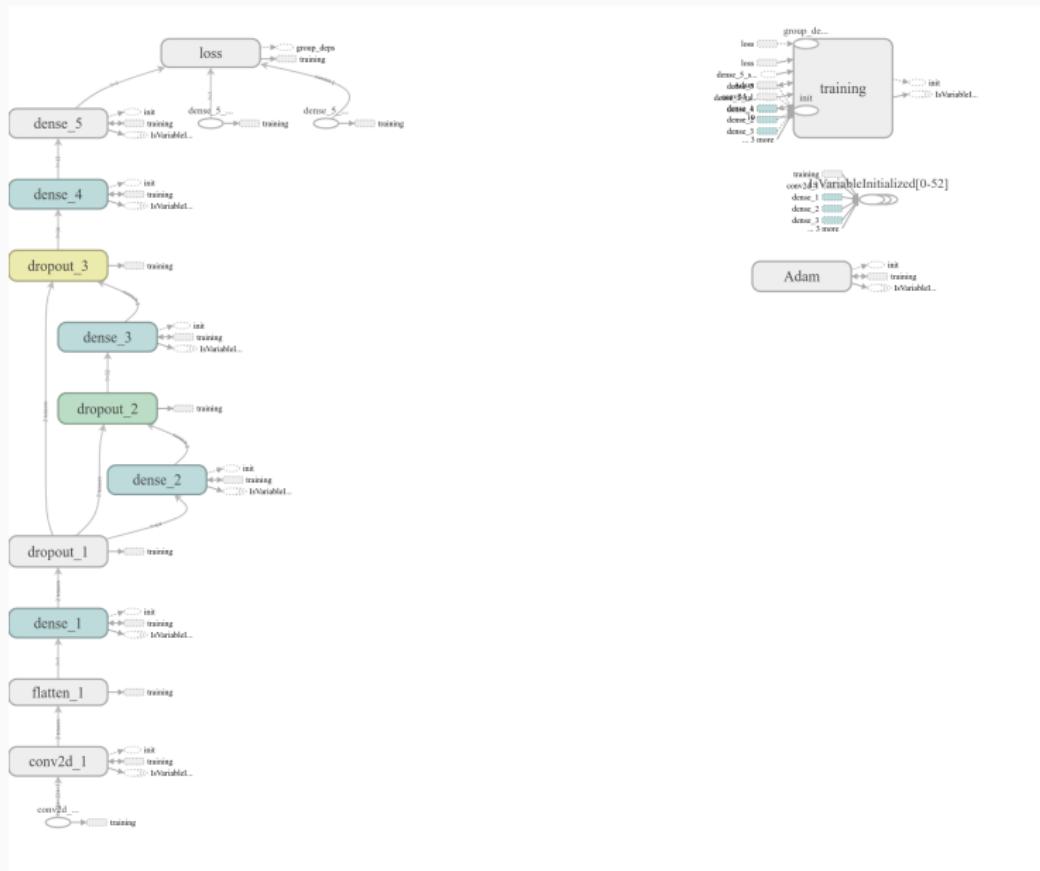
Se han implementado dos distintas arquitecturas de red neuronal con el propósito de comparar el rendimiento de una red neuronal convolucional comparado a una red tradicional.

Red neuronal tradicional

Capa	Tipo	Unidades (Filtros)	Dimensiones de salida	Parámetros entrenables	Función de Activación
input_1	InputLayer	-	(224,224,3)	0	
conv2d_1	Conv2D	3 (1 × 1)	(224,224,3)	12	ReLU
dense_1	Dense	64	64	9633856	ReLU
dropout_1	Dropout	-	64	0	-
dense_2	Dense	32	32	2080	ReLU
dropout_2	Dropout	-	32	0	-
dense_3	Dense	16	16	528	ReLU
dropout_3	Dropout	-	16	0	-
dense_4	Dense	12	12	204	ReLU
dense_5	Dense	1	1	13	Tanh

Total Capas	10
Total Parámetros	9636693

Red neuronal tradicional

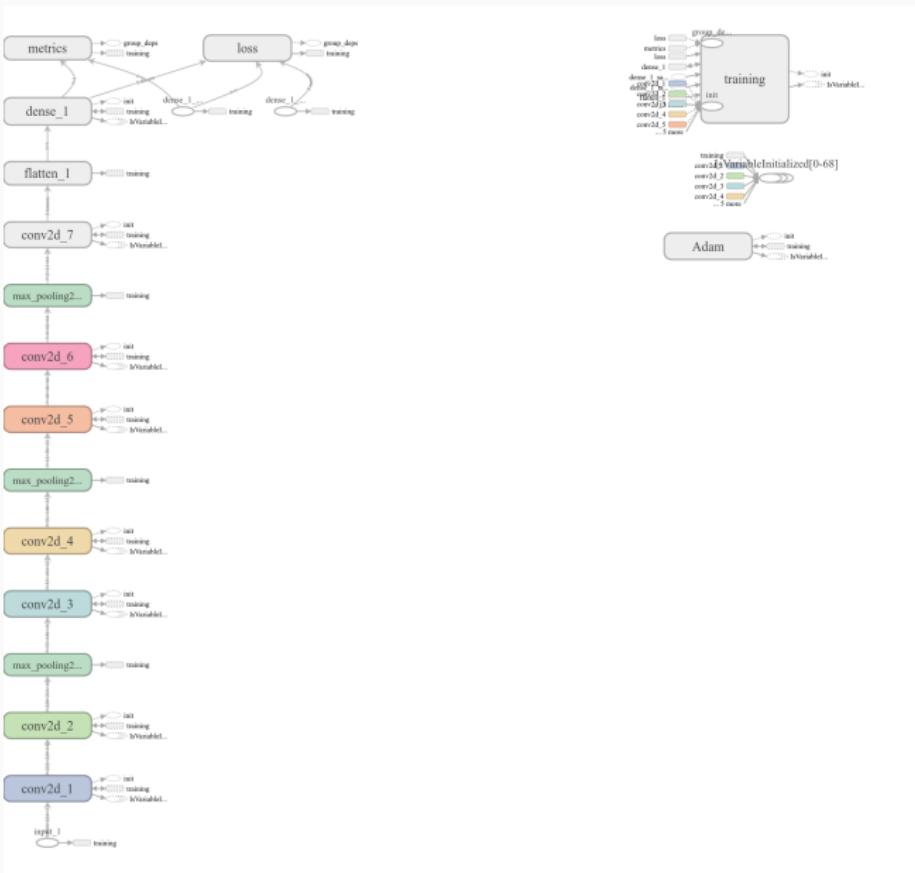


Red neuronal convolucional

Capa	Tipo	Unidades (Filtros)	Dimensiones de salida	Parámetros entrenables	Función de Activación
input_1	InputLayer	-	(224,224,3)	0	
conv2d_1	Conv2D	16 (3 × 5)	(222,220,16)	736	ReLU
conv2d_2	Conv2D	16 (3 × 5)	(220,216,16)	3856	ReLU
max_pooling2d_1	MaxPooling2D	-	(55,108,16)	-	-
conv2d_3	Conv2D	32 (3 × 5)	(53,104,32)	7712	ReLU
conv2d_4	Conv2D	32 (3 × 5)	(51,100,32)	15392	ReLU
max_pooling2d_2	MaxPooling2D	-	(12,50,32)	-	-
conv2d_5	Conv2D	64 (3 × 5)	(10,46,64)	30784	ReLU
conv2d_6	Conv2D	64 (3 × 5)	(8,42,64)	61504	ReLU
max_pooling2d_3	MaxPooling2D	-	(2,21,64)	-	-
conv2d_7	Conv2D	64 (3 × 5)	(2,21,4)	260	ReLU
dense_1	Dense	1	1	169	Tanh

Total Capas	12
Total Parámetros	120413

Red neuronal convolucional



Conjunto de datos

Balanceo del conjunto de datos

Debido a la naturaleza de las sesiones de entrenamiento, el conjunto de datos inicial se encuentra desbalanceado.

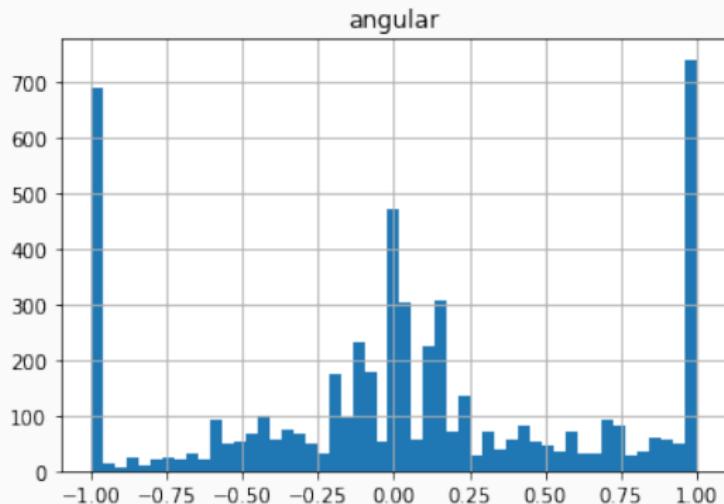


Figura 16: Distribución inicial del dataset. Fuente: Elaboración propia.

Balanceo del conjunto de datos

Se procede a realizar un balanceo del conjunto en base a dos tareas fundamentales:

- Eliminación de muestras con valor muy cercano o igual a cero.
- Aumentación de muestras con transformaciones aleatorias.

Balanceo del conjunto de datos

Una vez realizadas las operaciones se tiene un conjunto de datos con la siguiente distribución:

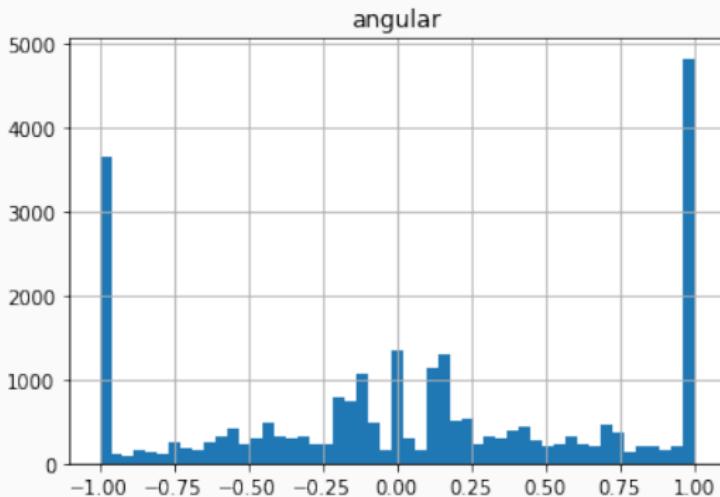


Figura 17: Distribución final del dataset. Fuente: Elaboración propia.

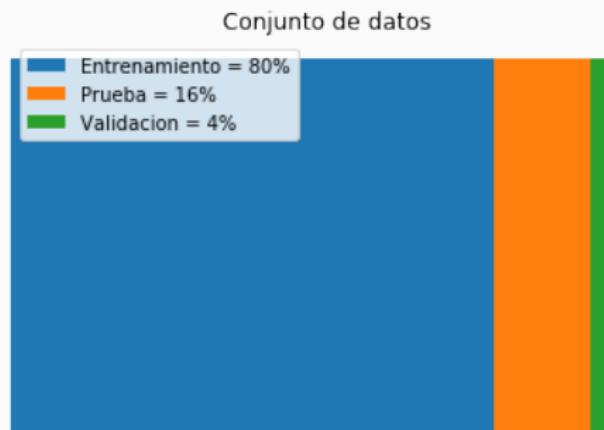
Separación de conjuntos de datos

Se ha separado el conjunto de datos en tres conjuntos con funciones distintas:

- **Entrenamiento:** Datos sobre los que se entrena la red neuronal en cada época.
- **Validación:** Pequeño conjunto sobre el cual se evalúa el error entre épocas.
- **Pruebas:** Conjunto no presente en el entrenamiento sobre el cual se realizan pruebas de rendimiento.

Separación de conjuntos de datos

Conjunto de datos	Muestras	Fracción
Entrenamiento	21084	80 %
Prueba	3691	16 %
Validación	1581	4 %
Total	26356	100 %



Análisis de Resultados

Entrenamiento

Error de Entrenamiento

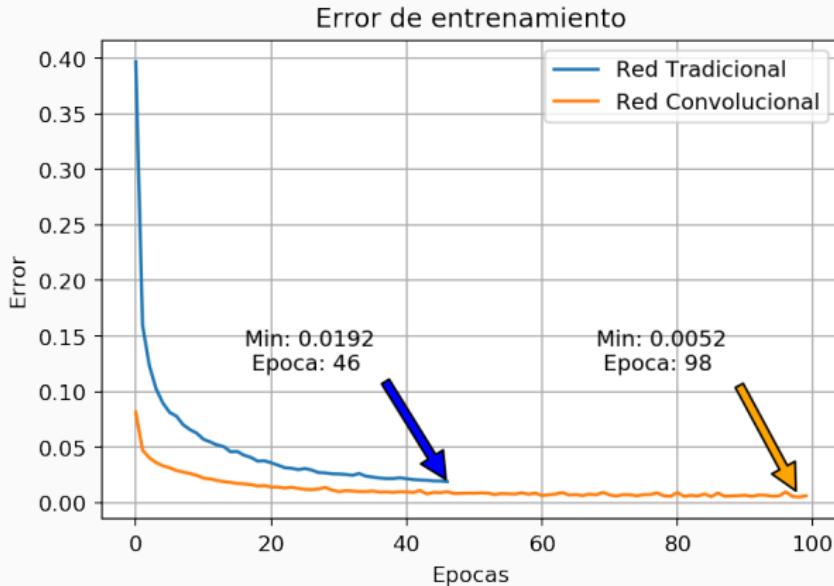


Figura 18: Error de entrenamiento. Fuente: Elaboración propia.

Error de Entrenamiento

- **El error de entrenamiento inicial es menor:** La capacidad de ajuste de parámetros es más eficaz en la red neuronal convolucional.
- **El error mínimo es menor:** La red convolucional se aproxima mejor al conjunto de entrenamiento.
- **Convergencia más rápida:** La velocidad de convergencia hacia el valor mínimo es casi el doble. La convergencia está directamente relacionada con el tiempo de entrenamiento requerido por cada arquitectura.

Error de Validación

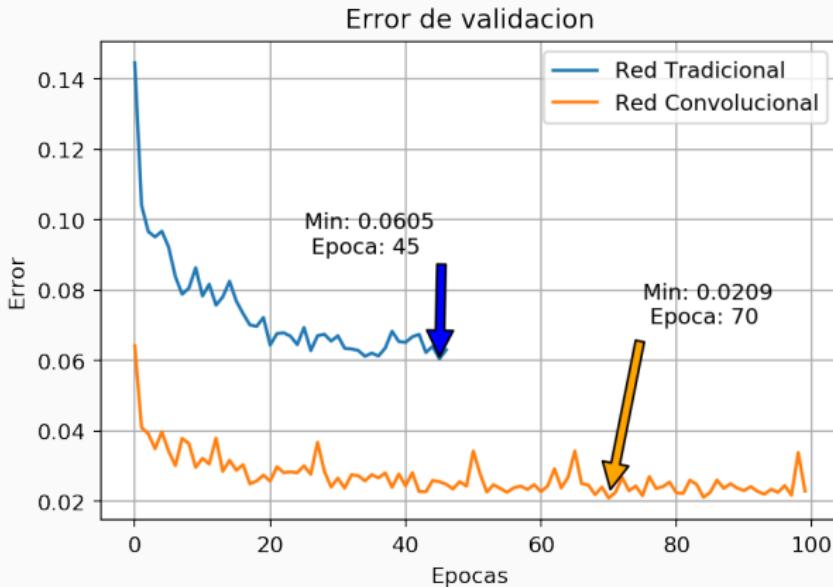


Figura 19: Error de validación. Fuente: Elaboración propia.

Error de Validación

- **El error de inicial es menor:** Con solamente una pasada por el conjunto de entrenamiento, se tiene una capacidad mayor de realizar predicciones cercanas a la realidad en casos nunca antes vistos.
- **El error mínimo es menor:** La capacidad de generalización de la red convolucional es mayor al de la red tradicional. Si el error de la red convolucional fuera mayor, esto podría indicar sobreentrenamiento de la red.
- **Convergencia:** Los valores del error son menores para la red convolucional mostrando que su rendimiento y capacidad de generalización son mayores al de una red tradicional.

Pruebas

Puntajes y errores en el conjunto de prueba

Los puntajes sobre el conjunto de prueba definen el rendimiento y la capacidad de generalización de la red sobre datos nunca antes vistos. Se usan las siguientes métricas:

- MSE: Error Cuadrático Medio.
- MAE: Error Absoluto Medio.
- R^2 : Coeficiente de Determinación.

Puntajes y errores en el conjunto de prueba

Modelo	MSE	MAE	R ²
Tradicional	0.0254	0.0976	0.9278
Convolucional	0.0160	0.0872	0.9484

Tabla 2: Evaluación de puntajes sobre el conjunto de prueba. Fuente:
Elaboración propia.

Puntajes y errores en el conjunto de prueba

Valor real	Predictión	Diferencia	Predictión de signo correcta
-0.297	-0.243	0.054	✓
1.0	0.921	0.079	✓
0.149	0.22	0.07	✓
-1.	-0.968	0.032	✓
-0.	0.069	0.069	✗

Tabla 3: Muestra de predicciones de la red convolucional. Fuente: Elaboración propia.

Pruebas de Campo

Pred: [-0.536]



Pred: [0.760]



Pred: [-0.855]



Pred: [-0.088]



Pred: [-0.997]



Pred: [1.0]



Figura 20: Ejemplos de predicción de la red convolucional. Fuente: Elaboración propia.

Representaciones Internas



Figura 21: Filtros de la penúltima capa convolucional. Fuente: Elaboración propia.

Representaciones Internas

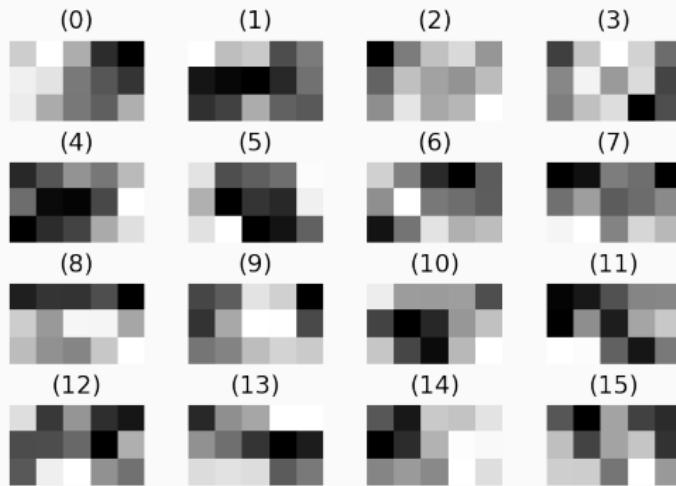
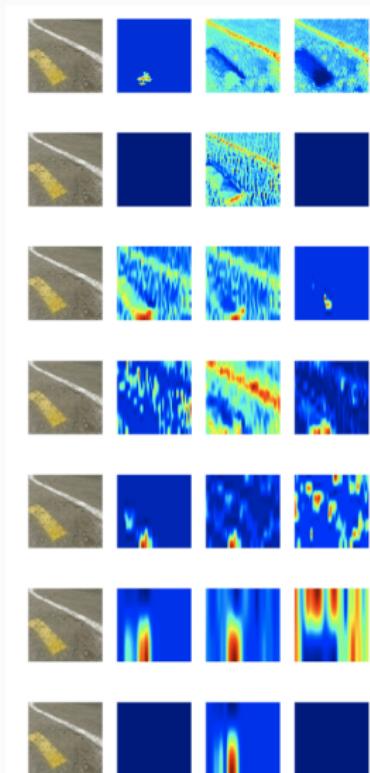


Figura 22: Filtros de la primera capa convolucional. Fuente: Elaboración propia.

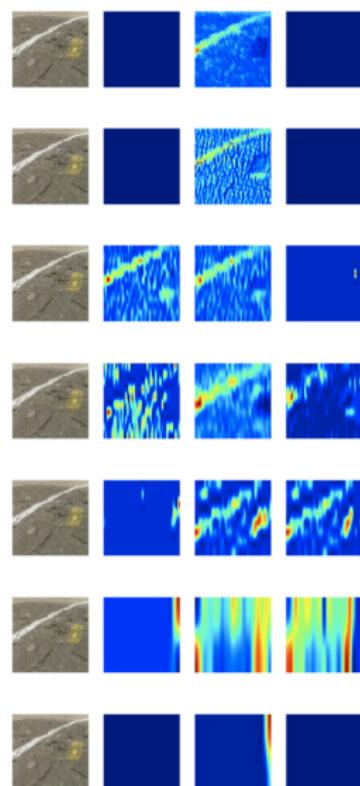
Representaciones Internas

Giro a la izquierda. Predicción: 1.



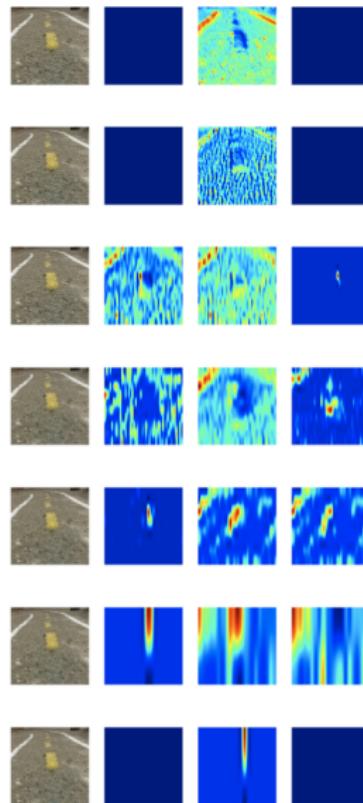
Representaciones Internas

Giro a la derecha. Predicción: **-0.84.**



Representaciones Internas

Conducción hacia adelante. Predicción: -0.1.



Conclusiones y Recomendaciones

Conclusiones i

- Se estudiaron los aspectos relacionados al desarrollo de sistemas de conducción autónoma, y posteriormente los conceptos básicos de aprendizaje automático, aprendizaje profundo, redes convolucionales y el proceso de entrenamiento de una red neuronal.
- Se ha introducido el esquema de la arquitectura de un sistema de conducción autónomo y se han planteado los requisitos y funcionalidades que debe tener el sistema en su conjunto y por cada subsistema y módulos correspondientes.
- Se han diseñado todos los subsistemas correspondientes en base a las limitaciones y herramientas de software y hardware elegidas. Los subsistemas son:
 - Subsistema de Control y Actuación.
 - Subsistema de Adquisición de Datos y Entrenamiento.
 - Subsistema de Inferencia y Control Autónomo

Conclusiones ii

- Se ha implementado una red neuronal convolucional para la tarea de la conducción autónoma.
- Se ha entrenado de manera satisfactoria una red neuronal convolucional.
- Se ha validado el rendimiento de la red neuronal convolucional para tareas de visión artificial en base a una comparación con otra arquitectura tradicional.

Recomendaciones i

- Generar **distintos conjuntos** de entrenamiento en condiciones climáticas diversas y ambientes variados.
- Diseñar, entrenar y validar arquitecturas de **redes neuronales distintas** o con variaciones a una red neuronal convolucional tradicional.
- Extender la implementación de la red a una tarea de **clasificación categórica**.
- Implementar el sistema desarrollado en **otros modelos de vehículos**, como por ejemplo, tracción diferencial.
- Extender el sistema presentado en este proyecto con **tareas avanzadas**, incluyendo en el flujo de trabajo, módulos de planificación de trayectorias, detección de objetos y programación de misiones.

Recomendaciones ii

- Crear una **línea de investigación** dedicada al diseño e implementación de sistemas robóticos inteligentes y vehículos autónomos en el IEA.
- Considerar este proyecto y sus resultados como una **plataforma de desarrollo** sobre la cual se puedan implementar diversos algoritmos y aplicaciones en sistemas robóticos inteligentes y vehículos autónomos.

Preguntas?

Referencias i

-  "Linear regression," Nov 2018. [Online]. Available: https://en.wikipedia.org/wiki/Linear_regression
-  R. Szeliski, *Computer Vision*. Springer-Verlag GmbH, 2011. [Online]. Available: https://www.ebook.de/de/product/9630757/richard_szeliski_computer_vision.html
-  C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York Inc., 2006. [Online]. Available: https://www.ebook.de/de/product/5324937/christopher_m_bishop_pattern_recognition_and_machine_learning.html
-  D. Cornelisse, "An intuitive guide to convolutional neural networks," Apr 2018. [Online]. Available: <https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050>

Referencias ii

-  "Nucleo-f303k8." [Online]. Available:
<https://www.st.com/en/evaluation-tools/nucleo-f303k8.html>
-  T.-y. Wang, "From perceptron to deep learning," Jan 2016. [Online]. Available: <http://databeauty.com/blog/2018/01/16/From-Perceptron-to-Deep-Learning.html>
-  D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.