UNIVERSIDAD
NACIONAL
DE COLOMBIA

Convolutional Neural Networks (CNN)

2023 – 01

Ph.D. JOHN W. BRANCH

Profesor Titular
Departamento de Ciencias de la Computación y de la Decisión
Director del Grupo de I+D en Inteligencia Artificial – GIDIA
jwbranch@unal.edu.co

M.Sc. CAMILO LAITON

Ingeniero de Visión por Computadora
Allen Institute for Neural Dynamics
claiton@unal.edu.co – camilo.Laiton@alleninstitute.org

Contenido

1. Introducción
2. Funcionamiento de la capa de convolución
    1. Filtros
    2. Mapas de características
3. Capas de pooling
4. Arquitectura de las redes neuronales convolucionales
5. Aplicaciones en el campo de las imágenes

Origen de la capa de convolución

La capa de convolución tiene su origen del estudio de la corteza visual del cerebro de los gatos.

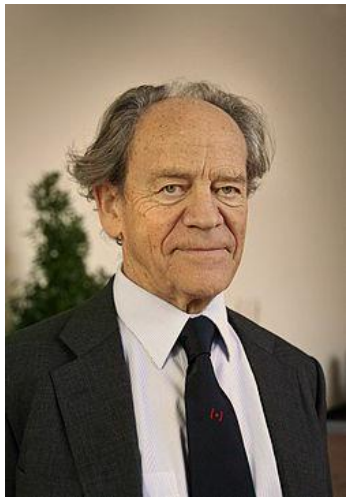


Fig 1. David Hubel y Torsten Wiesel.

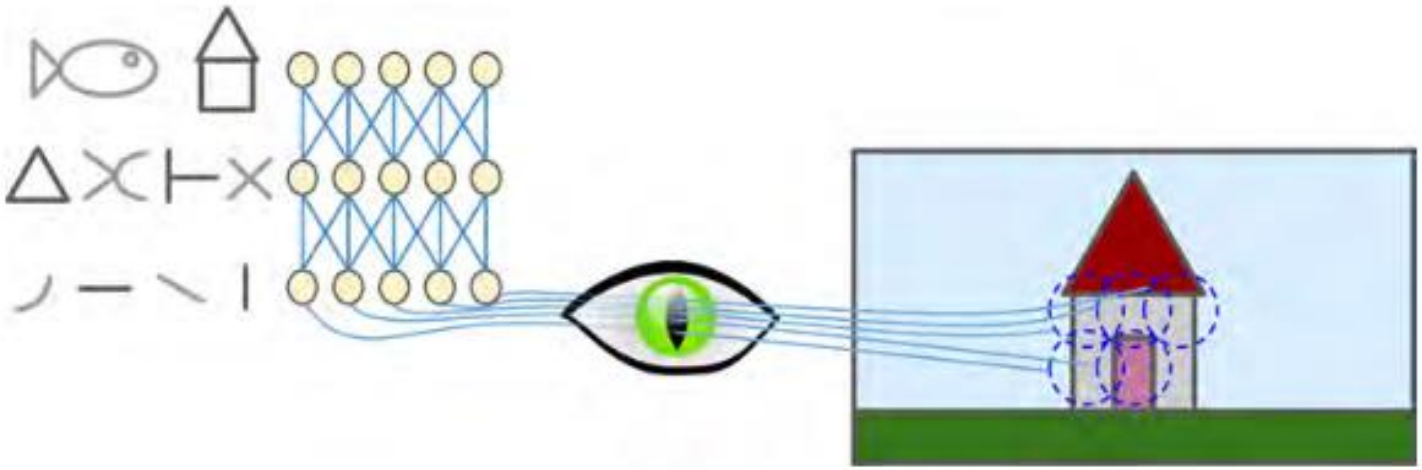


Fig 2. Diagrama simplificado del funcionamiento de los campos receptivos locales en el ojo de los gatos.

[1] Citar hands on machine learning

Origen de la capa de convolución

Las CNNs tienen su origen desde el neocognitrón [] en los años 80s. Sin embargo, su primer gran éxito en el campo de la computación fue en 1998 con el articulo *Gradient-based learning applied to document recognition* [].

Los autores propusieron un modelo computacional con:

- Aprendizaje a partir de los datos.
- Función de optimización.
- Capas sucesivas de convoluciones.

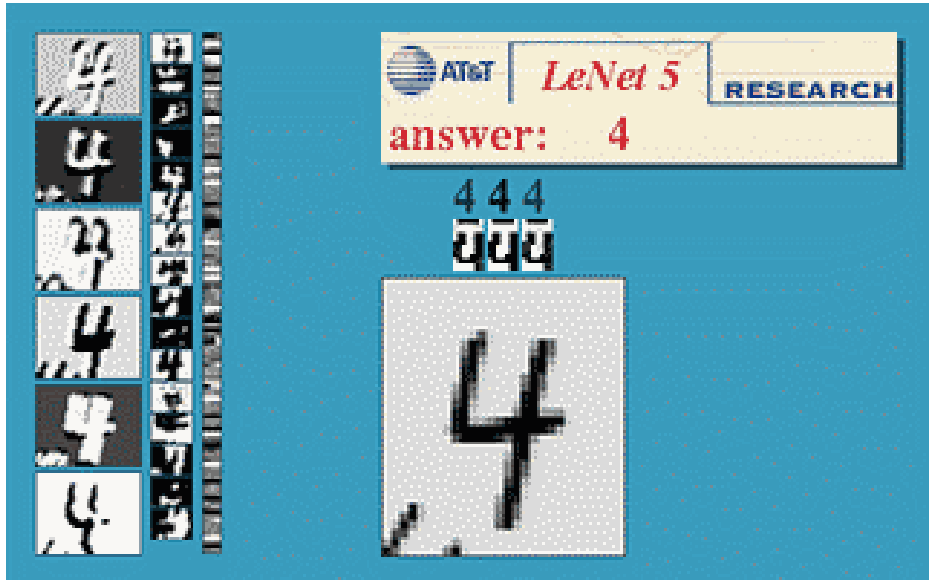


Fig 3. Primer gran éxito de las CNNs aplicada al reconocimiento de dígitos.

[1] Citar neocognitron https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf y document recognition lecun http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

Redes neuronales convolucionales (CNN)

Las CNN son una combinación de varias ideas:

1. Campos receptivos locales (“extracción” de características elementales visuales).
2. Pesos en su arquitectura (coeficientes entrenables llamados filtros).
3. Sub-sampling temporal o espacial (reducción del tamaño original de los datos).

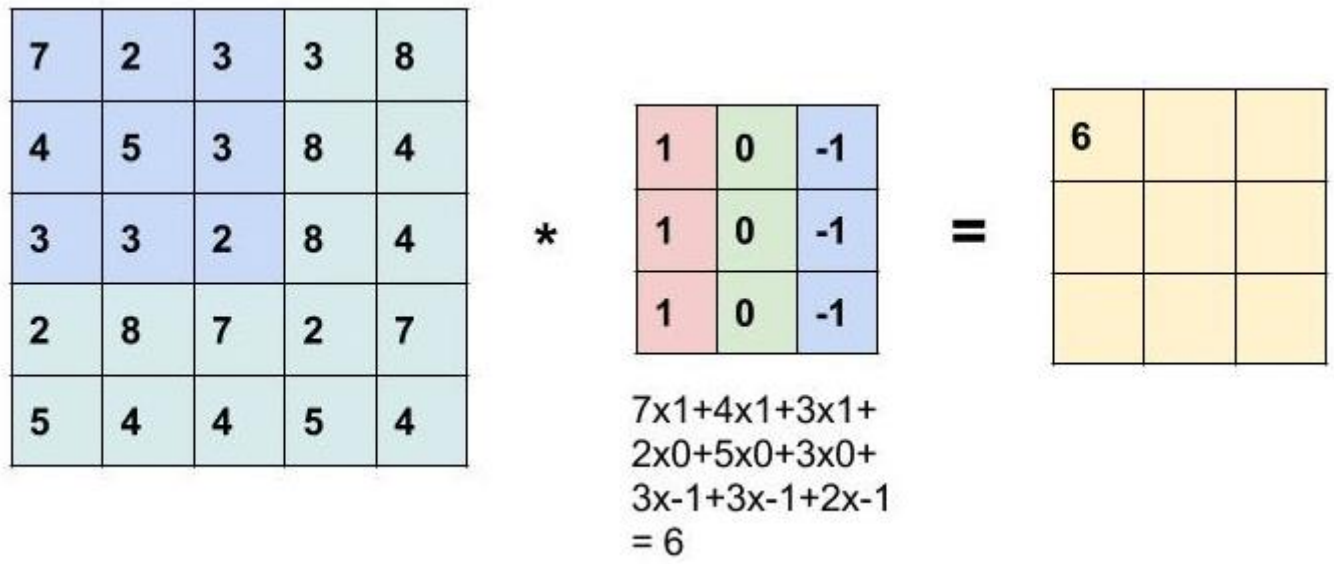


Fig 4. Funcionamiento de una CNN.

Redes neuronales convolucionales (CNN)

Los filtros (también llamados como *filtro convolutivo)* son los pesos entrenables de una capa de convolución. Este representa nuestro *campo receptivo local.*

Los filtros son un *stack* de kernels y tienen las siguientes características:

- Tienen un tamaño fijo definido por el usuario.
- Son entrenables.
- Capaz de detectar distintos tipos de características del campo receptivo local.
- Se mueven en la imagen basado en un parámetro llamado *stride.*

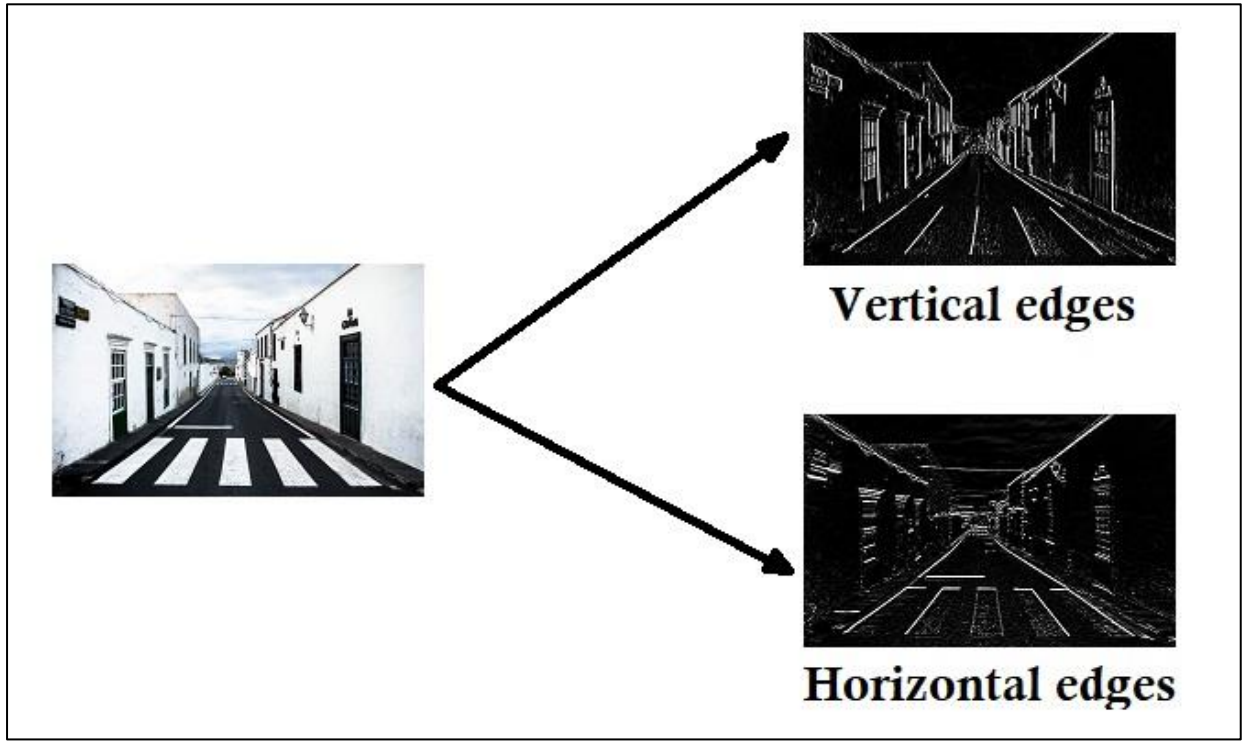


Fig 5. Ejemplo de kernel para extracción de bordes verticales y horizontales.

Capas de agrupación

Las capas de agrupación (también llamadas *pooling*) nos permiten superar las limitaciones de memoria en el entrenamiento de nuestras arquitecturas basadas en redes convolutivas.

Los *pooling* más famosos son:

- **Max Pooling:** Solo deja el máximo valor de la grilla.
- **Average Pooling**: Promedio de los valores de la grilla.

Fig 6. Imagen luego de aplicar Max Pooling.

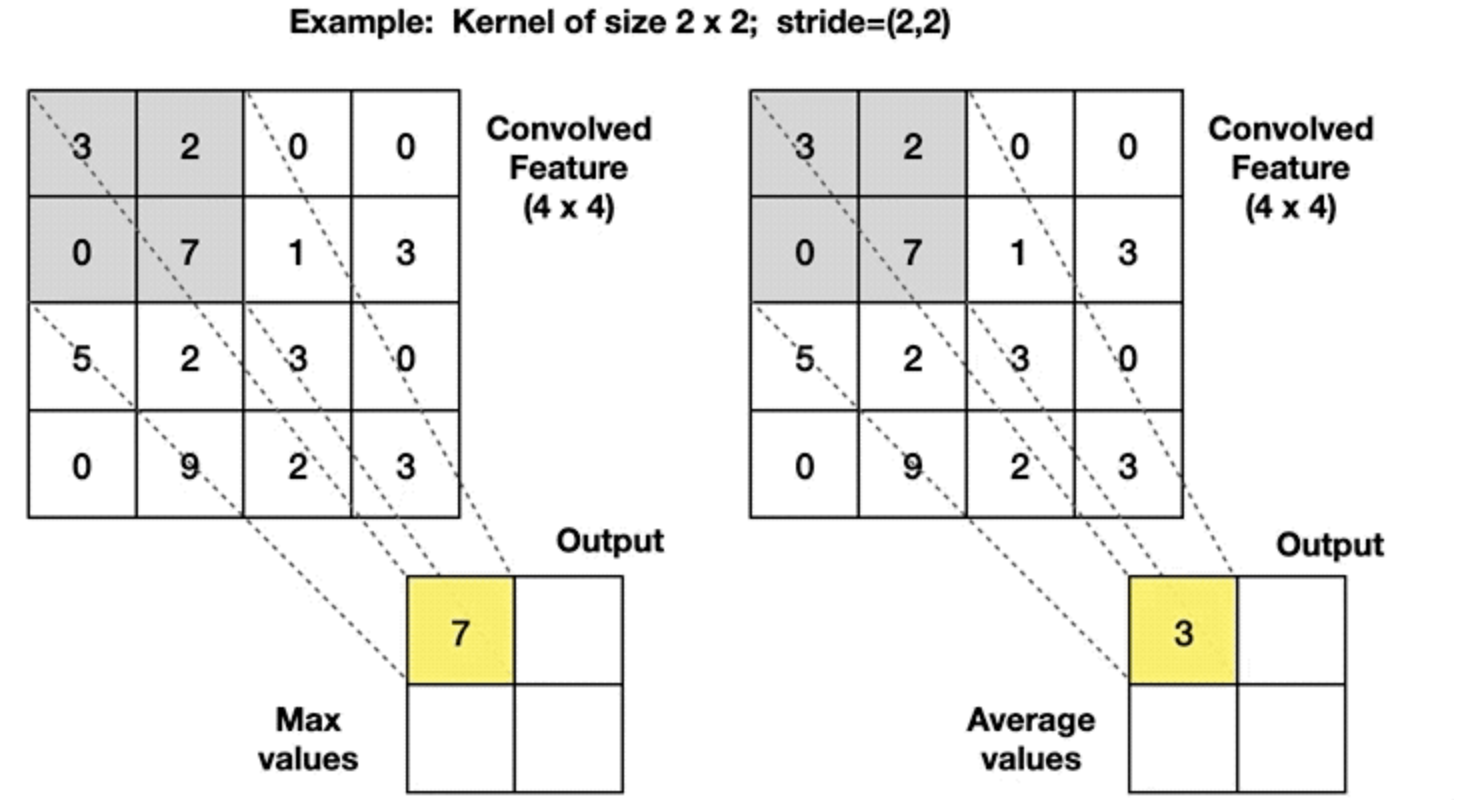
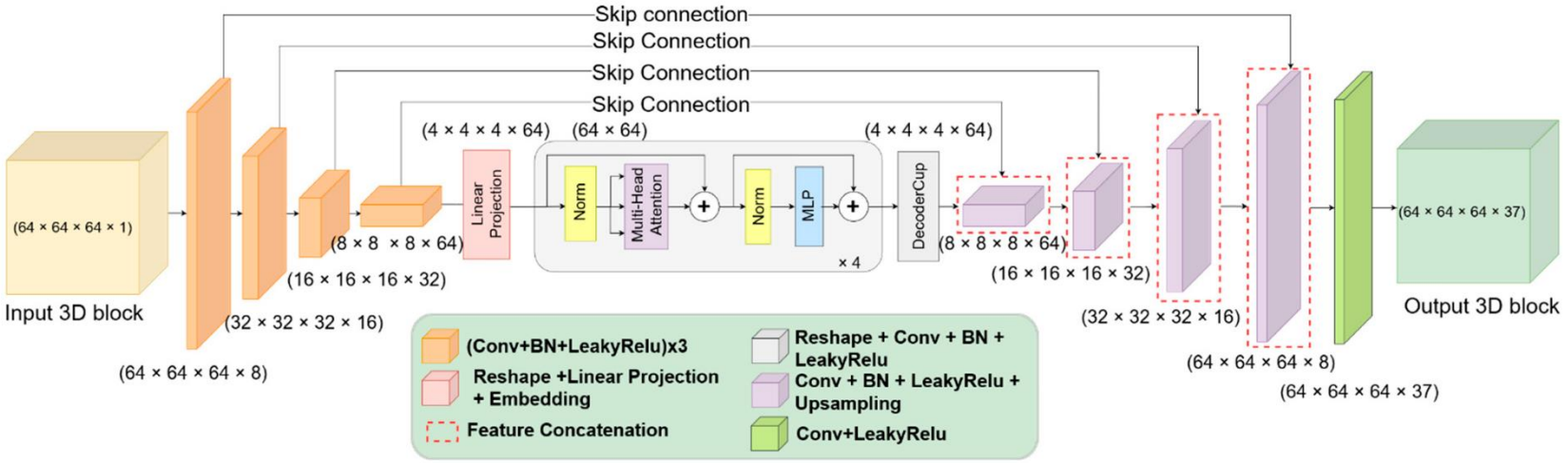


Fig 7. Operaciones de agrupación.

Capas sucesivas de convoluciones

El poder de las CNNs radica en la extracción de características de alto orden en donde los mapas de características obtenidos en la capa anterior son utilizados en la convolución de capas posteriores.



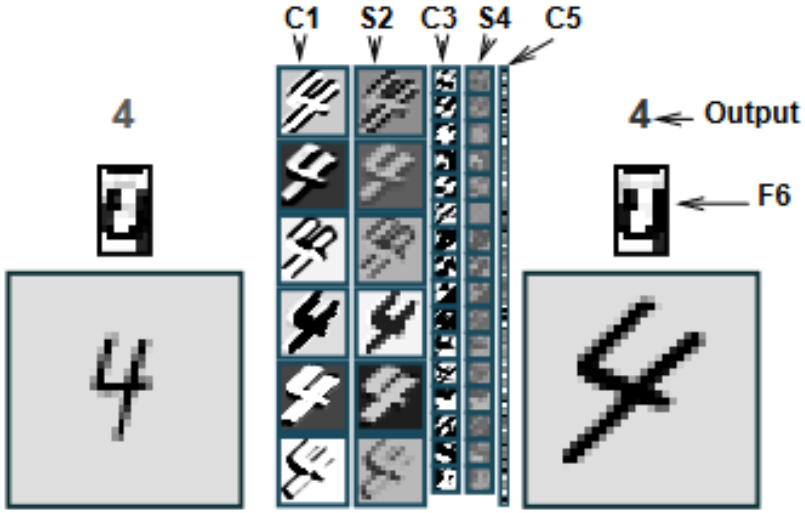


Fig 8. Ejemplos de arquitecturas de redes neuronales basadas en convolución tomadas de [1] y [3].

[3] C. Laiton-Bonadiez, G. Sanchez-Torres, and J. Branch-Bedoya, "Deep 3D Neural Network for Brain Structures Segmentation Using Self-Attention Modules in MRI Images," Sensors, vol. 22, no. 7, 2022, doi: 10.3390/s22072559.

Limitaciones de las CNN

Ejemplo en limitaciones de memoria para una capa de CNN:

Parámetros:

- Tamaño del kernel: 3 x 3
- Número de filtros: 100
- Entrada de imagen: 256 x 256
- Modelo de color: RGB
- Profundidad de bit: 16

Número de parámetros:

$((3x3)x(3)+1)x100 = 2800$

Número de multiplicaciones flotantes:

$(100)x(256x256)x((3x3)x3) = 176\,947\,200$

Aproximadamente **177 millones** de multiplicaciones flotantes!

Memoria requerida:

$(100)x(256x256)x(32) = 209\,715\,200$ bits equivalente a **26.2 MB en memoria RAM**.

Limitaciones de las CNN

Requieren grandes cantidades de datos y memoria para entrenarse.	Perdida de características de la imagen por operaciones de pooling [4].
Requiere de grandes cantidades de datos para entrenarse [4].	Incapacidad de reconocer la pose, textura y deformaciones de la imagen [5].

[4] M. Kwabena Patrick, A. Felix Adekoya, A. Abra Mighty, y B. Y. Edward, "Capsule Networks – A survey", Journal of King Saud University - Computer and Information Sciences, sep. 2019, doi: 10.1016/j.jksuci.2019.09.014.

[5] S. Sabour, N. Frosst, y G. E. Hinton, "Dynamic Routing Between Capsules", CoRR, vol. abs/1710.09829, 2017, [En línea]. Disponible en: http://arxiv.org/abs/1710.09829

Aplicaciones – Imágenes médicas

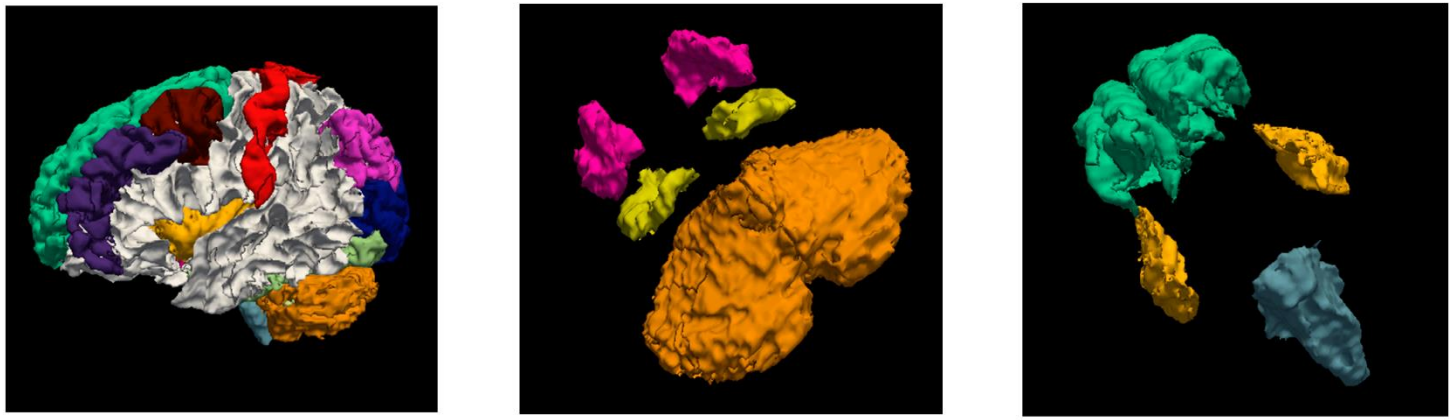


Fig 9. Ejemplo de segmentación de estructuras cerebrales tomada de [3].

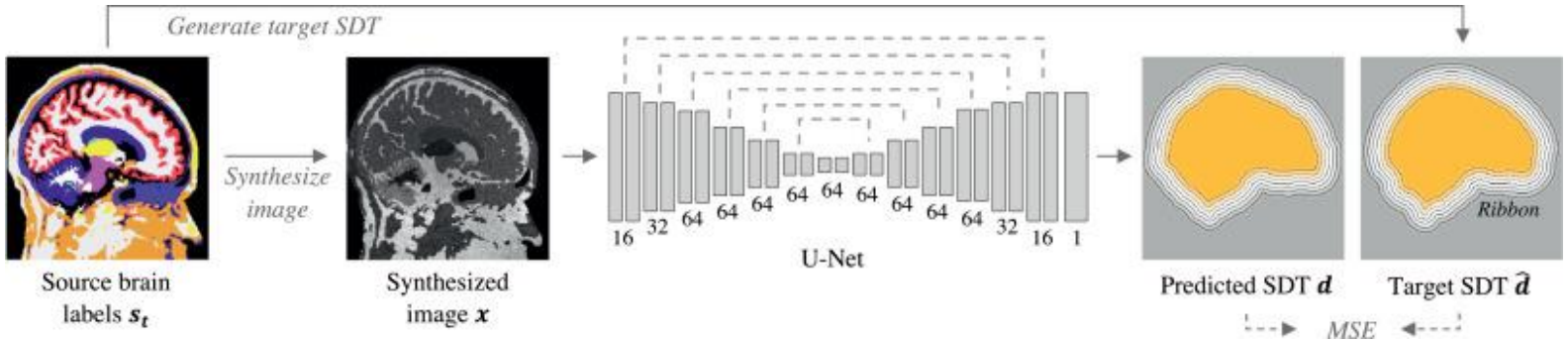
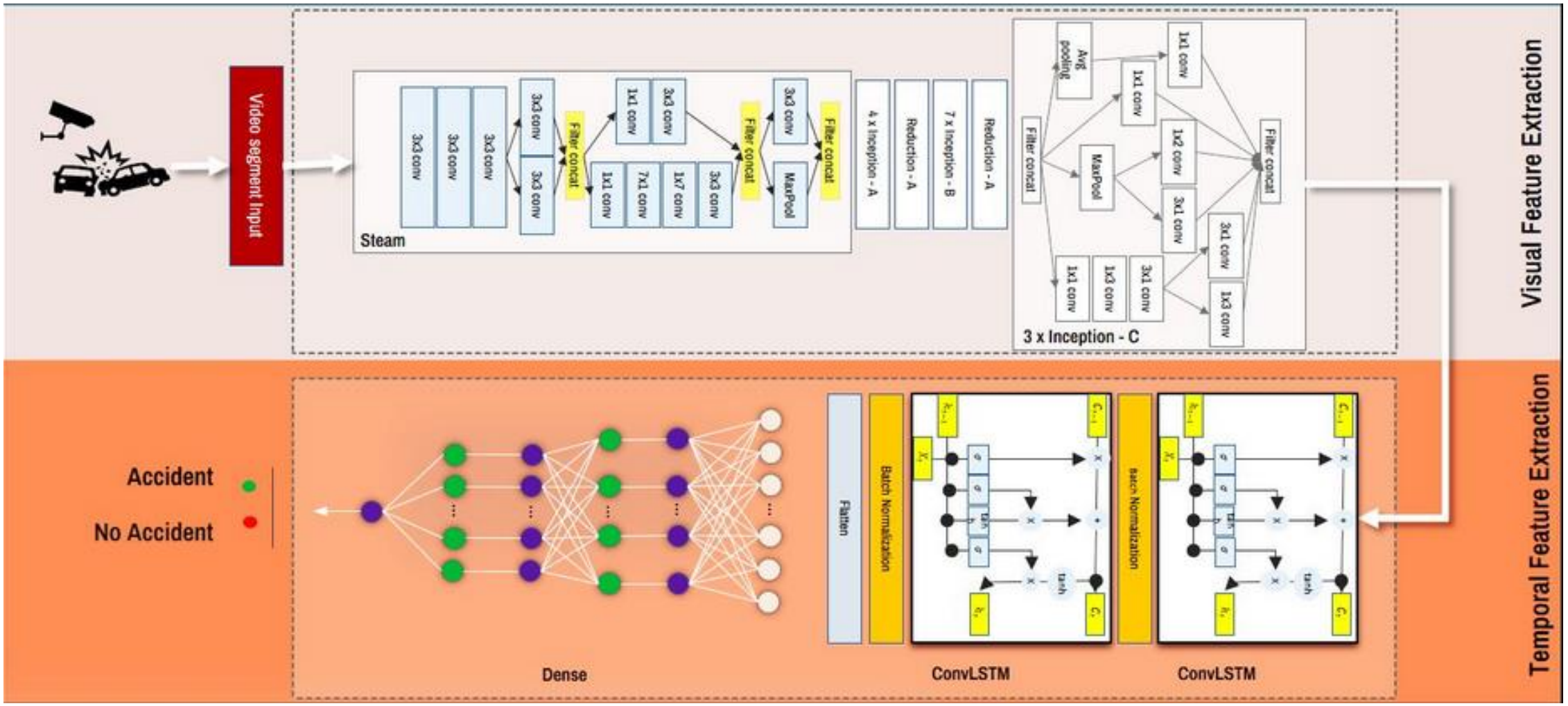


Fig 9. Ejemplo de extracción de tejido cerebrale tomada de [6].

https://www.sciencedirect.com/science/article/pii/S1053811922005900

Aplicaciones – Clasificación de accidentes de tráfico



Citar tesis de sergio

Gidia Grupo de I+D en Inteligencia Artificial

UNIVERSIDAD NACIONAL DE COLOMBIA

Preguntas