

CNN y visión por computadora en autos autónomos

Presented by

Chiquete Martínez Karen Daniela

“¿Cómo sabe un auto autónomo que eso que ve a 30 metros no es una bolsa, sino un niño corriendo?”



La Inteligencia Artificial (IA), especialmente las Redes Neuronales Convolucionales (CNN), es el “sistema visual” del auto.

¿Qué es la Visión por Computadora en un Auto Autónomo?

La visión por computadora (computer vision) en un vehículo autónomo es el componente del sistema de percepción que utiliza **cámaras** y **algoritmos** para “**ver**” y entender el entorno



Tareas principales de Visión

Applications and Challenges of Computer Vision in Autonomous Driving

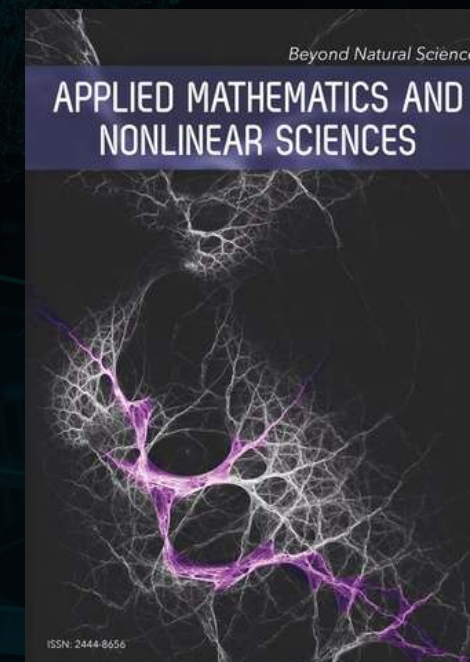
“ La visión por computadora permite la detección de entidades de tráfico, la segmentación de imágenes y el reconocimiento de objetos críticos mediante **algoritmos de aprendizaje profundo** (como redes neuronales convolucionales) para mejorar la seguridad y la inteligencia del vehículo ”

► **Detección de objetos**

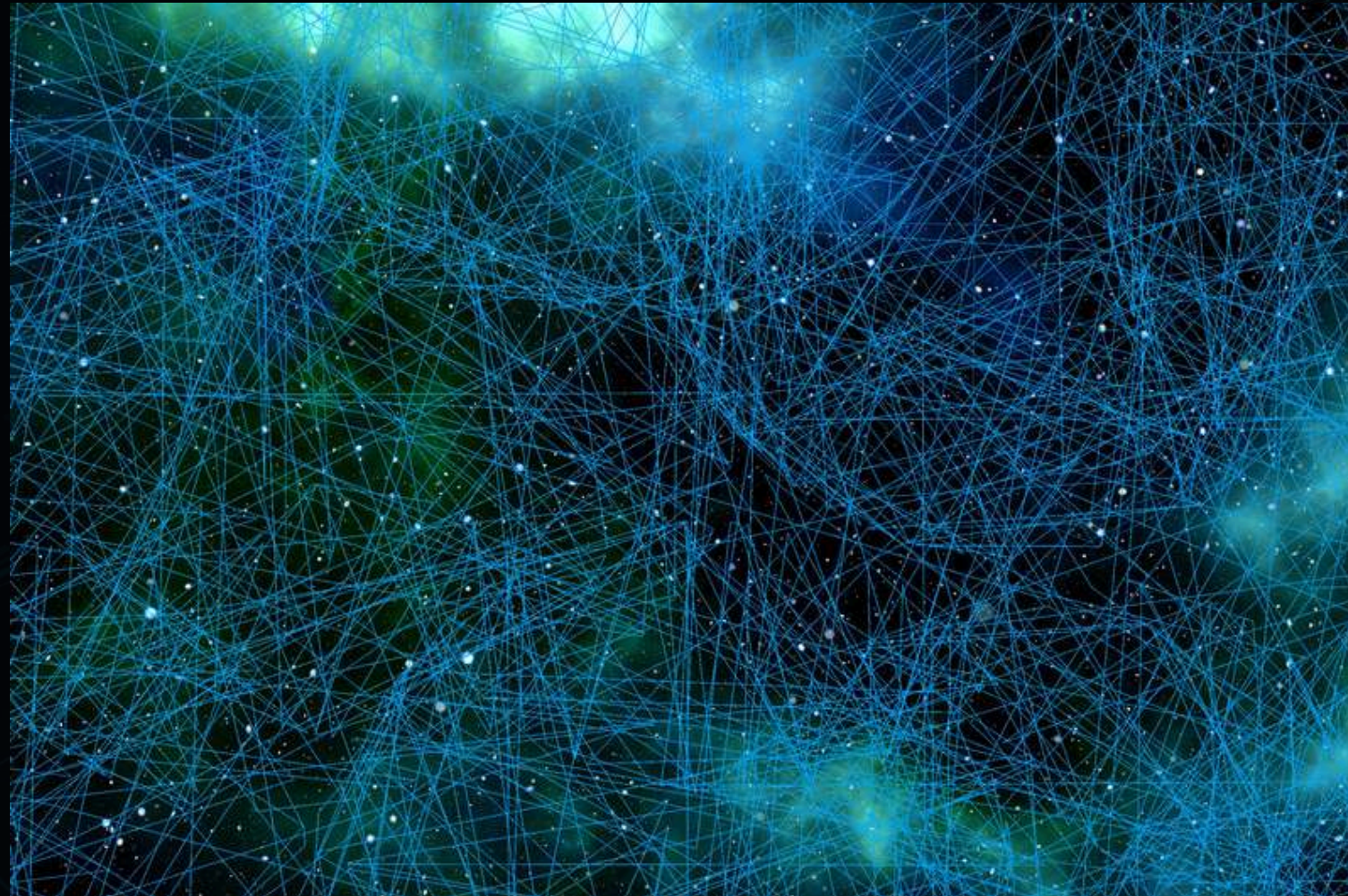
► **Estimación de distancia**

► **Segmentación del entorno**

► **Seguimiento de objetos**



Liu, Jiahao, and Peng Ren. "Applications and Challenges of Computer Vision in Autonomous Driving" *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Erdos Publish House S.L., 2024, <https://doi.org/10.2478/amns-2024-1549>



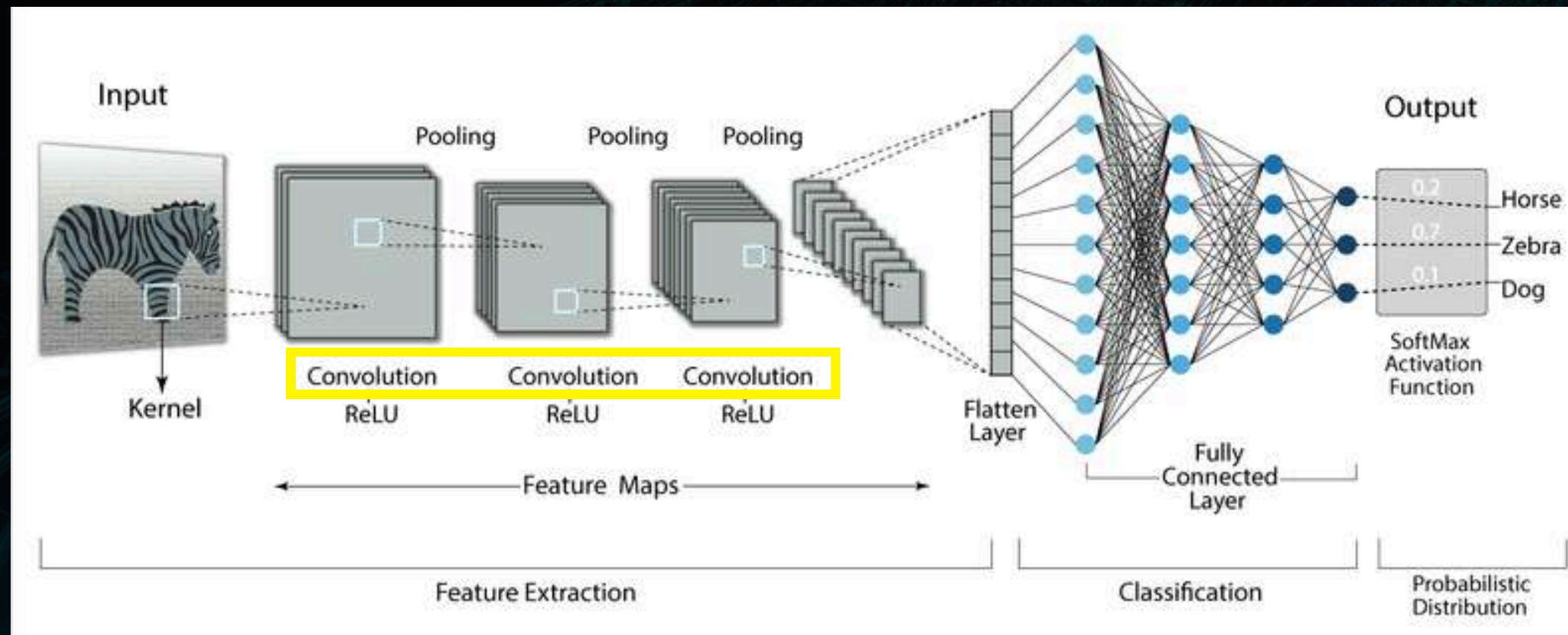
¿Qué es una CNN y por qué revolucionó la conducción autónoma?



CNN

Una CNN (Red Neuronal Convolutiva) es un tipo de red neuronal profunda especialmente diseñada para procesar imágenes (y otros datos con estructura espacial). En lugar de mirar toda la imagen de golpe, una CNN “**barre**” la imagen con pequeños filtros (**kernels**) que extraen características locales, y luego construye a partir de eso representaciones cada vez más complejas.

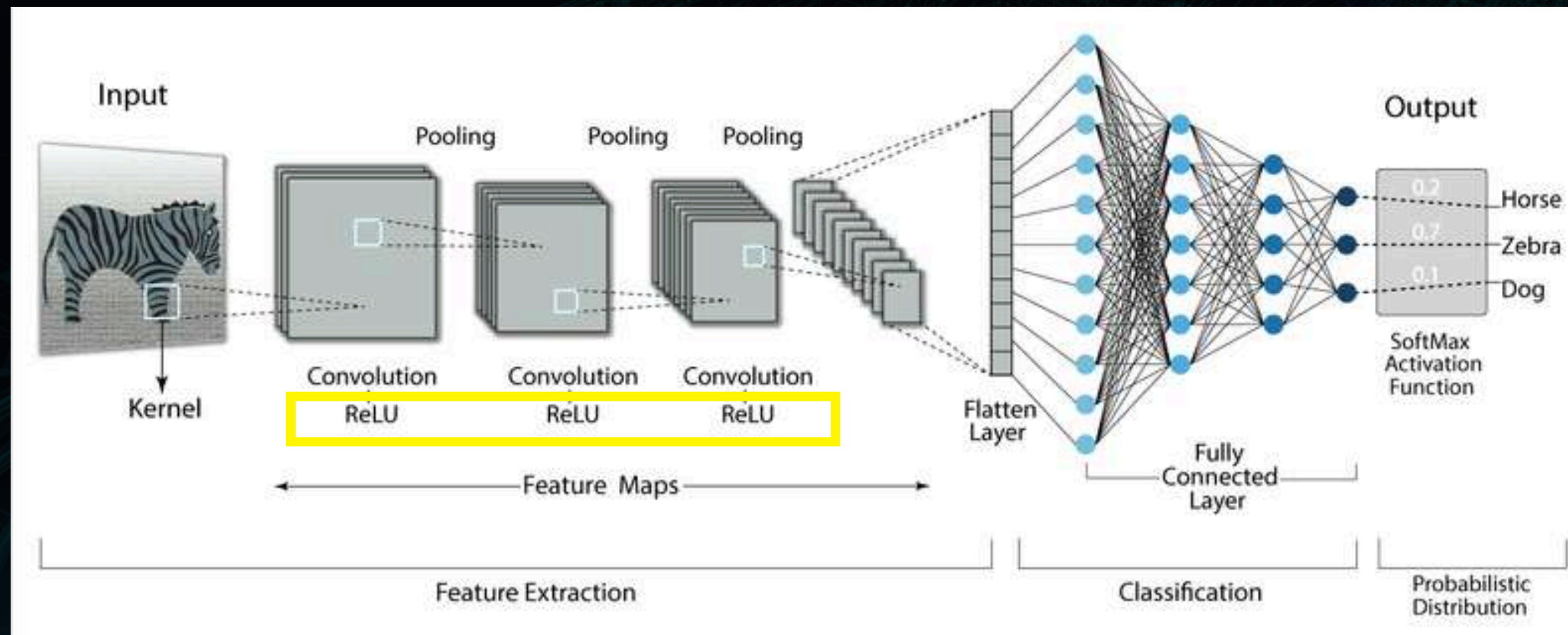
Componentes clave de una CNN



»» Convoluciones

La operación central: un filtro (matriz pequeña) se desliza ("barre") sobre la imagen, multiplicando valores, para obtener un mapa de características. Permite detectar localmente patrones como bordes, texturas, esquinas.

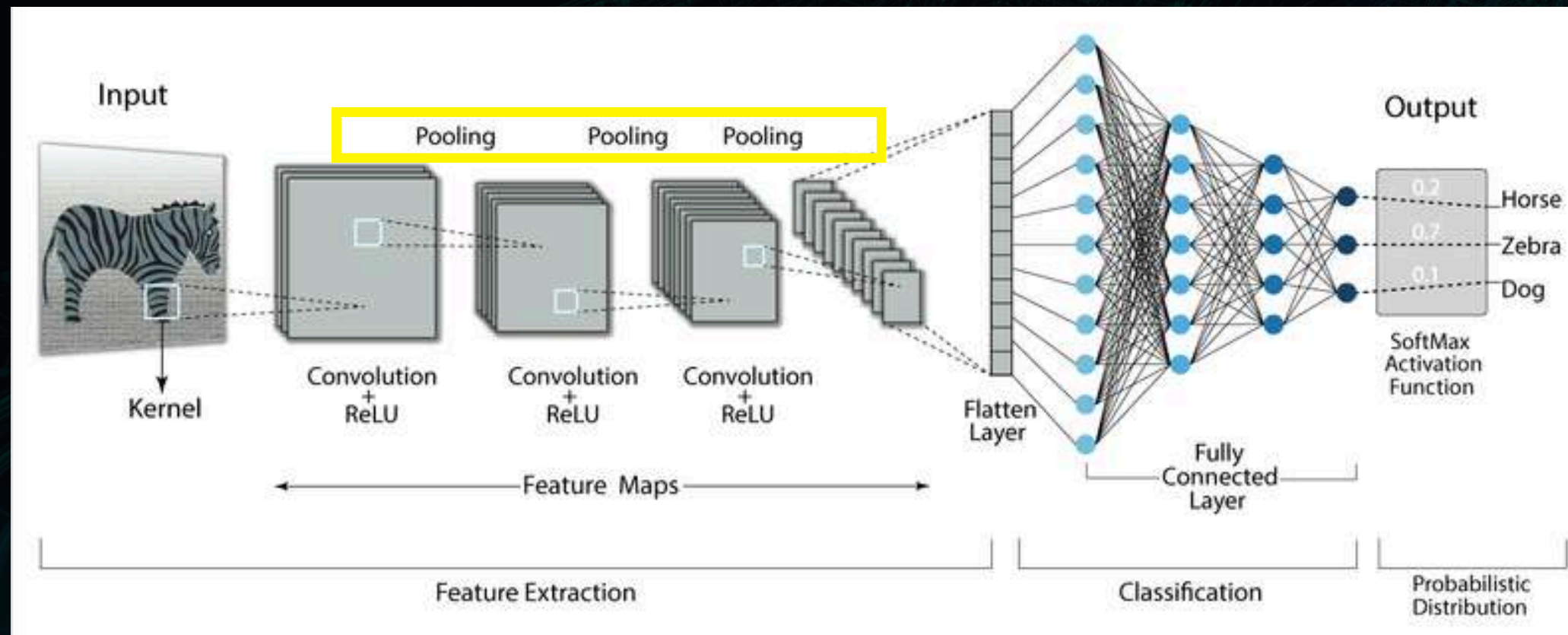
Componentes clave de una CNN



» ReLU (Rectified Linear Unit)

Función de activación no lineal: convierte valores negativos a cero, mantiene valores positivos. Hace que la red aprenda patrones no triviales sin volver el modelo demasiado complejo.

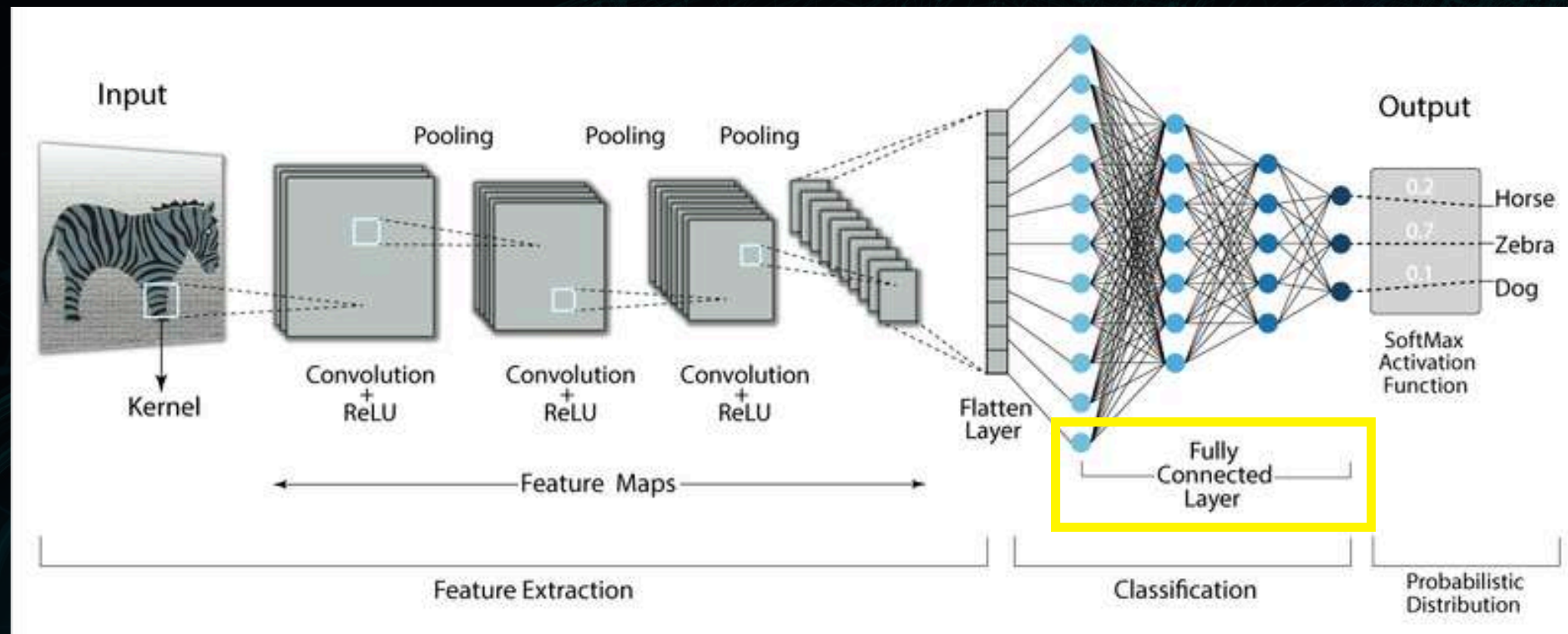
Componentes clave de una CNN



» Pooling

Se usa para reducir el tamaño espacial de los mapas de características (ej. max pooling toma el valor máximo en una ventana). Ayuda a hacer la red más eficiente, menos sensible a pequeños cambios y más rápida.

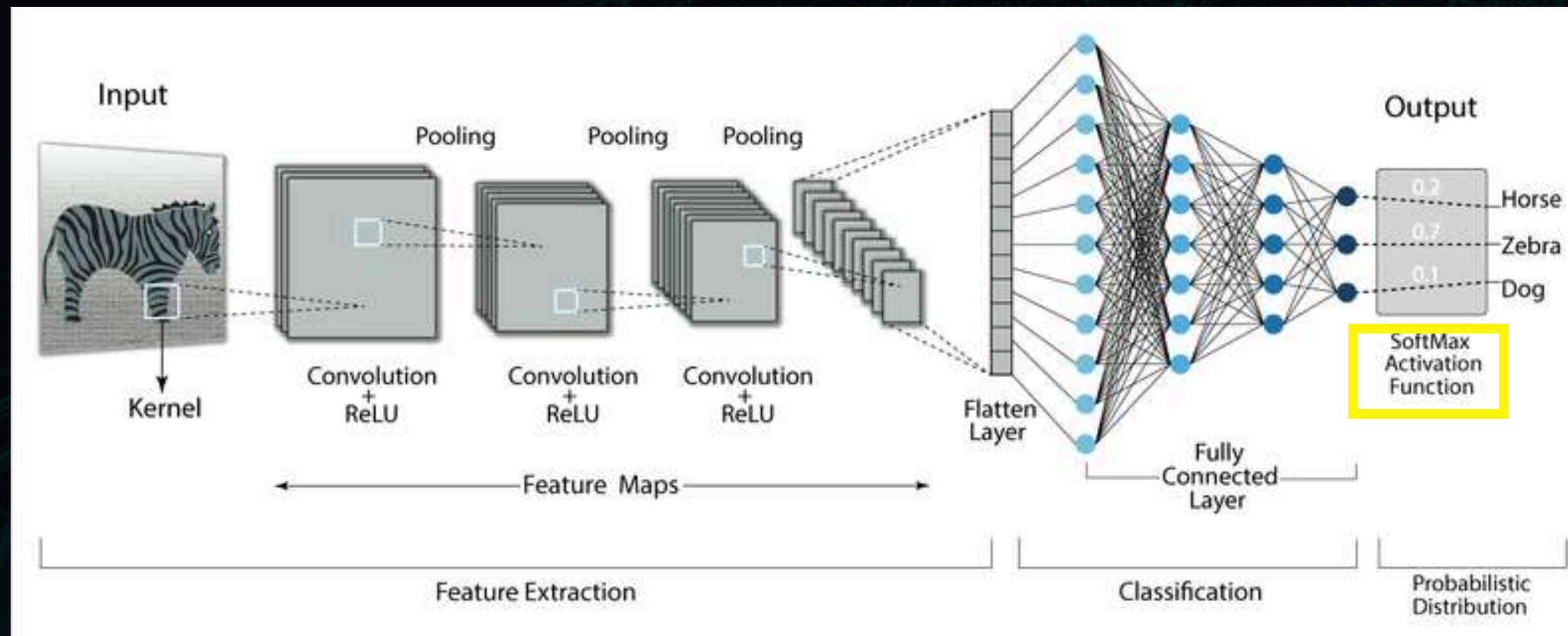
Componentes clave de una CNN



» Fully connected (capas completamente conectadas)

En las capas finales, las neuronas se conectan a todas las de la capa anterior, como en una red neuronal tradicional. Aquí es donde se combinan todas las características abstractas aprendidas para hacer la predicción final (clasificación, decisión, etc.).

Componentes clave de una CNN



» Softmax

Función usada típicamente en la capa de salida para clasificación: convierte los valores en probabilidades (suma 1). Permite que la red “decida” qué clase es más probable (por ejemplo: “auto”, “peatón”, “señal”).

¿Por qué funcionó tan bien para autos autónomos?

▶ Aprenden directamente de millones de imágenes reales

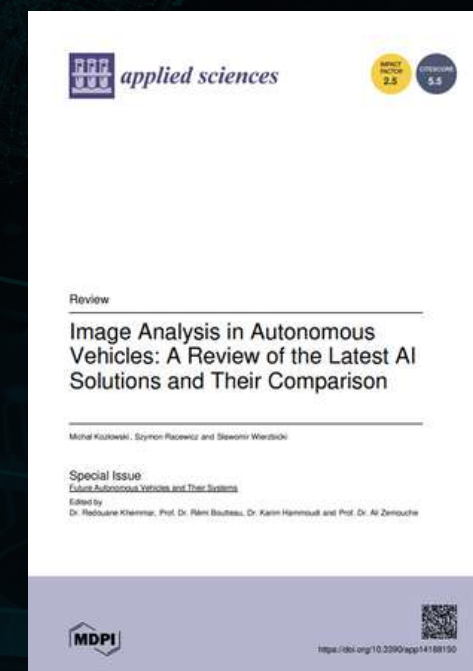
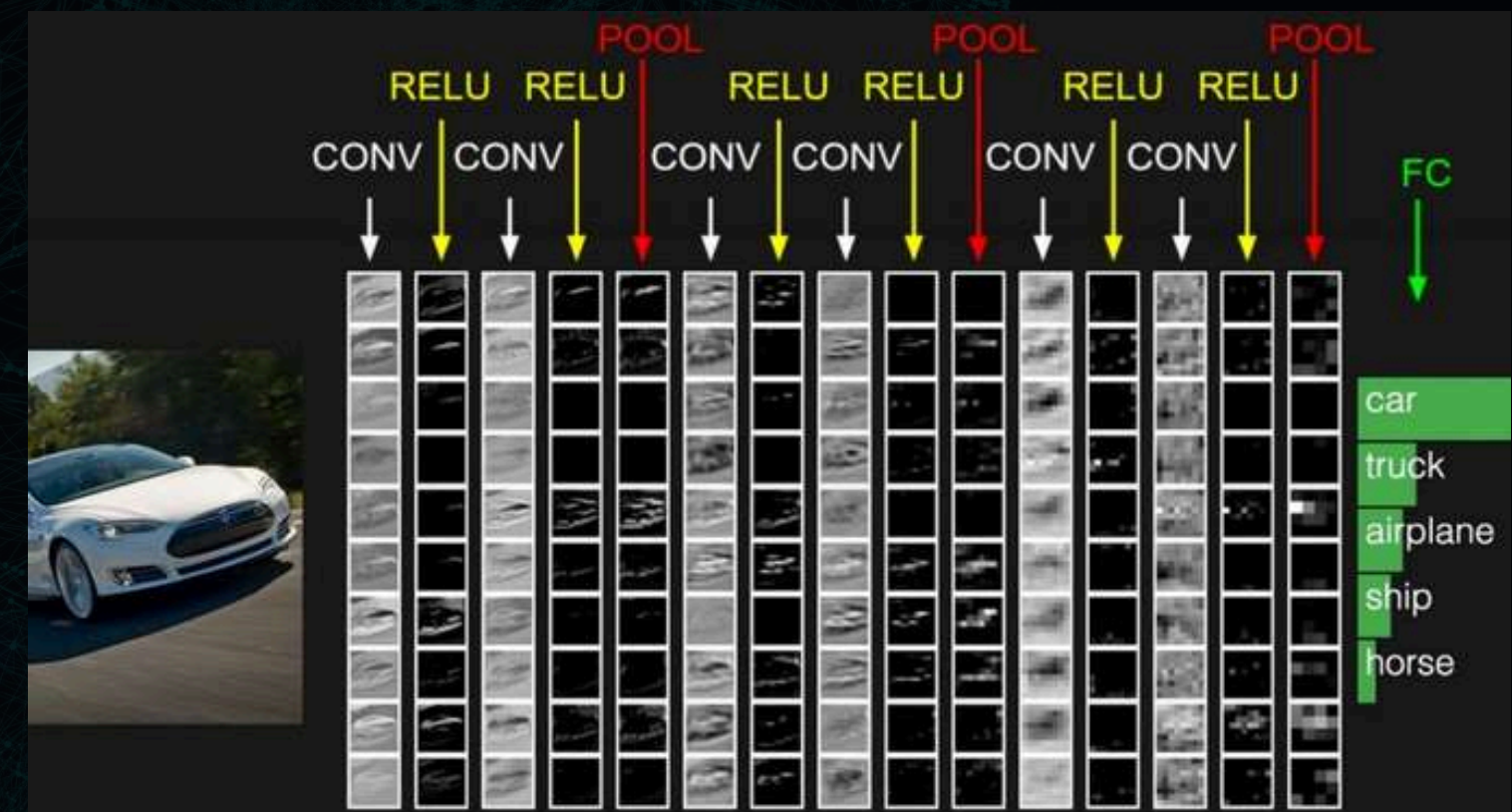
En la conducción autónoma, los datos visuales son enormes (cámaras, video, escenarios urbanos). Las CNN pueden entrenarse con grandes **datasets** (como Cityscapes, nuScenes) para aprender directamente de imágenes reales.

▶ Objetos

▶ Señales

▶ Peatones

▶ Carriles



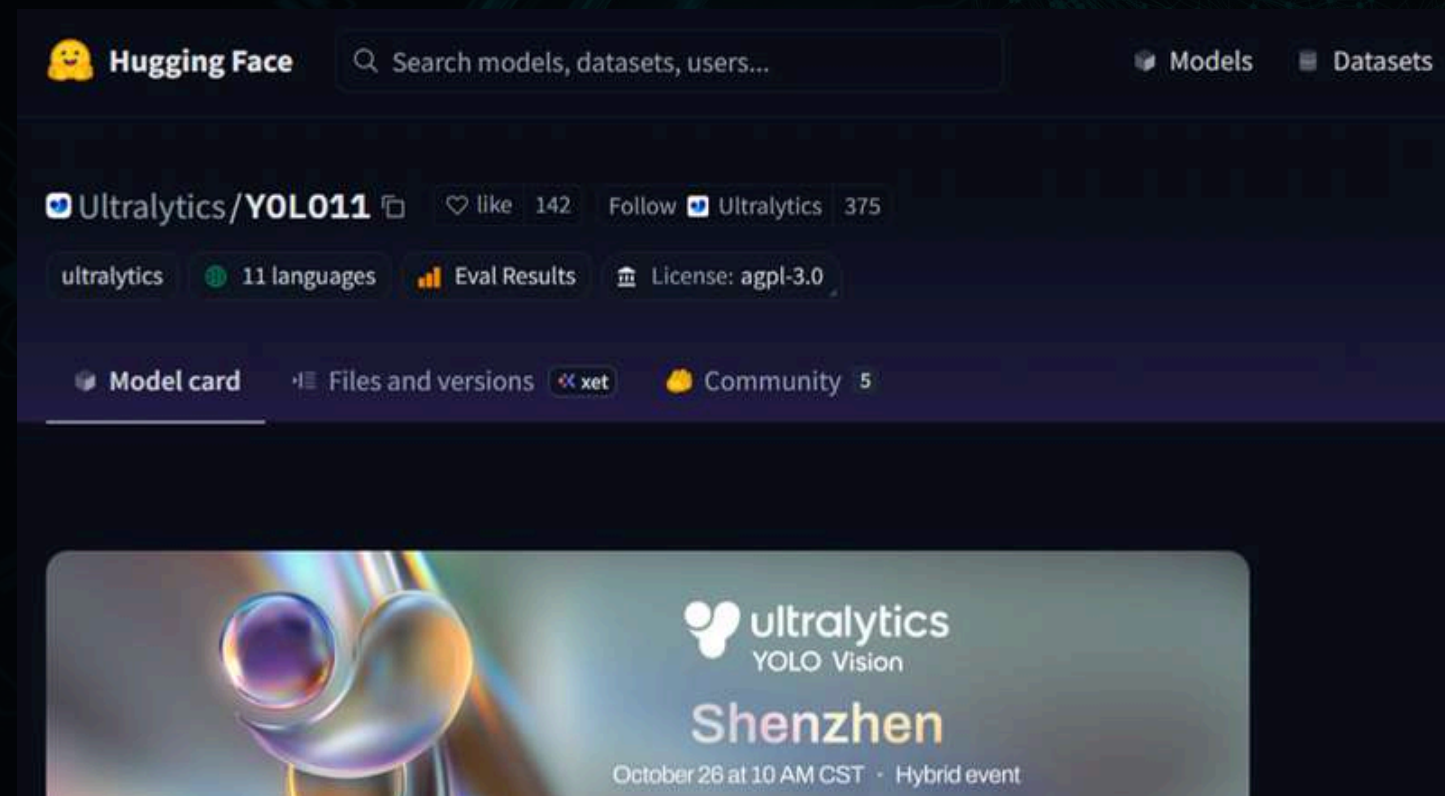
Kozłowski, M., Racewicz, S., & Wierzbicki, S. (2024). Image Analysis in Autonomous Vehicles: A Review of the Latest AI Solutions and Their Comparison. *Applied Sciences*, 14(18), 8150.

<https://doi.org/10.3390/app14188150>

Cómo las CNN permiten que un auto “entienda” su entorno

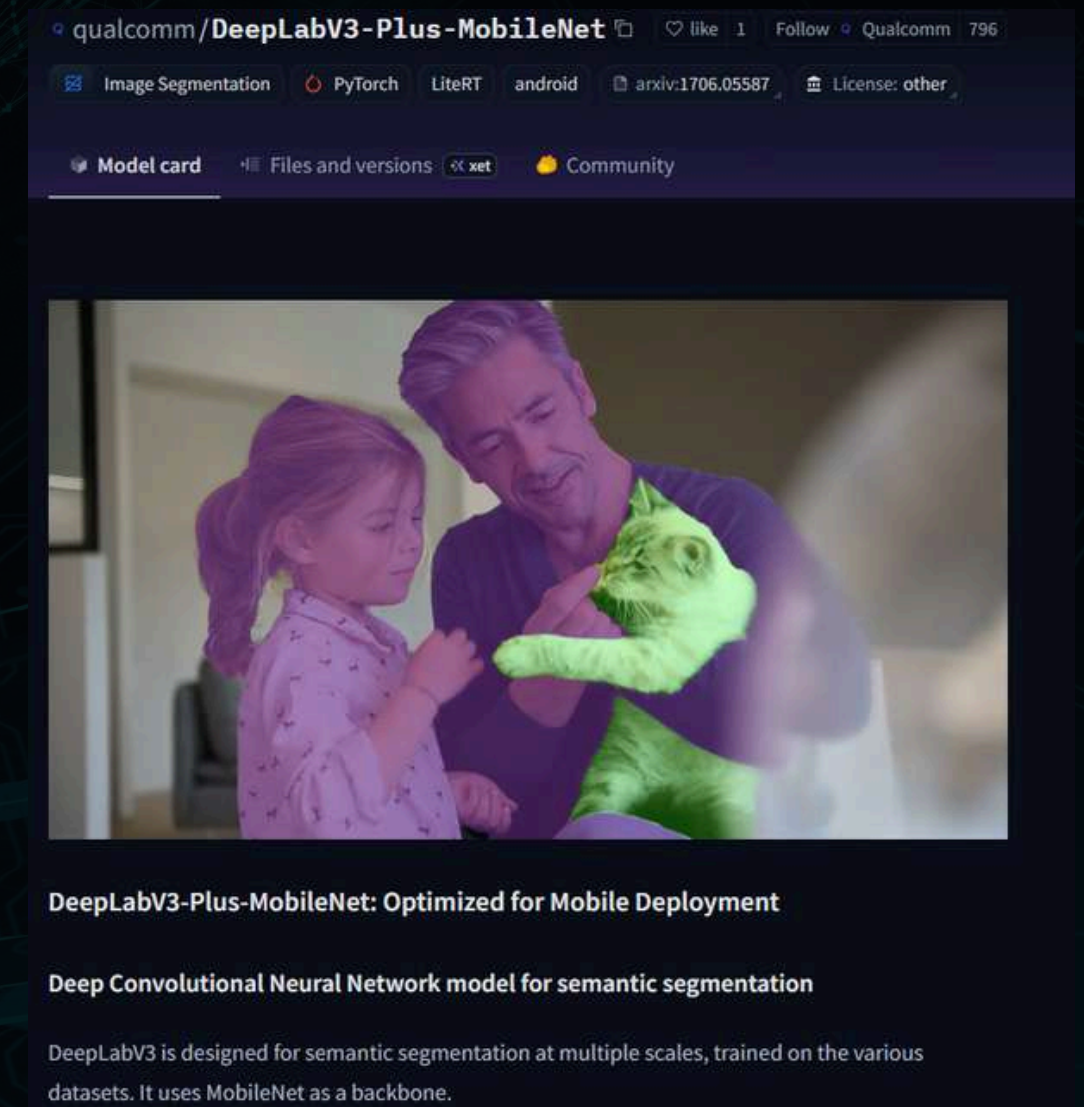
Detección de Objetos

Las CNN permiten que el auto identifique y localice objetos importantes dentro de la escena, como autos, peatones, bicicletas, semáforos o animales. A través de modelos como **YOLO** o **Faster R-CNN**, el vehículo puede colocar “cajas” alrededor de cada elemento y reconocerlo en milisegundos. Esta detección rápida y precisa es esencial para evitar colisiones y entender qué ocurre delante, detrás y a los lados del auto.



Segmentación del Entorno

Mediante técnicas de **segmentación semántica**, las CNN analizan cada píxel de la imagen para determinar si pertenece a carretera, acera, césped, edificios u obstáculos. Esto permite que el auto identifique con precisión por dónde puede circular y qué zonas debe evitar. Modelos como **DeepLab** o **SegNet** ayudan a crear un mapa visual detallado del entorno, incluso en calles complejas o con elementos parcialmente visibles.





Seguimiento y Predicción de Movimiento

Una vez que las CNN detectan personas, autos u otros elementos, el sistema de IA utiliza modelos de seguimiento para determinar su trayectoria. Esto permite saber hacia dónde se mueven los peatones, si un auto está cambiando de carril o si un ciclista va a cruzar. Gracias a esta predicción, el vehículo puede anticiparse a situaciones de riesgo y ajustar su velocidad o dirección antes de que ocurra un accidente.



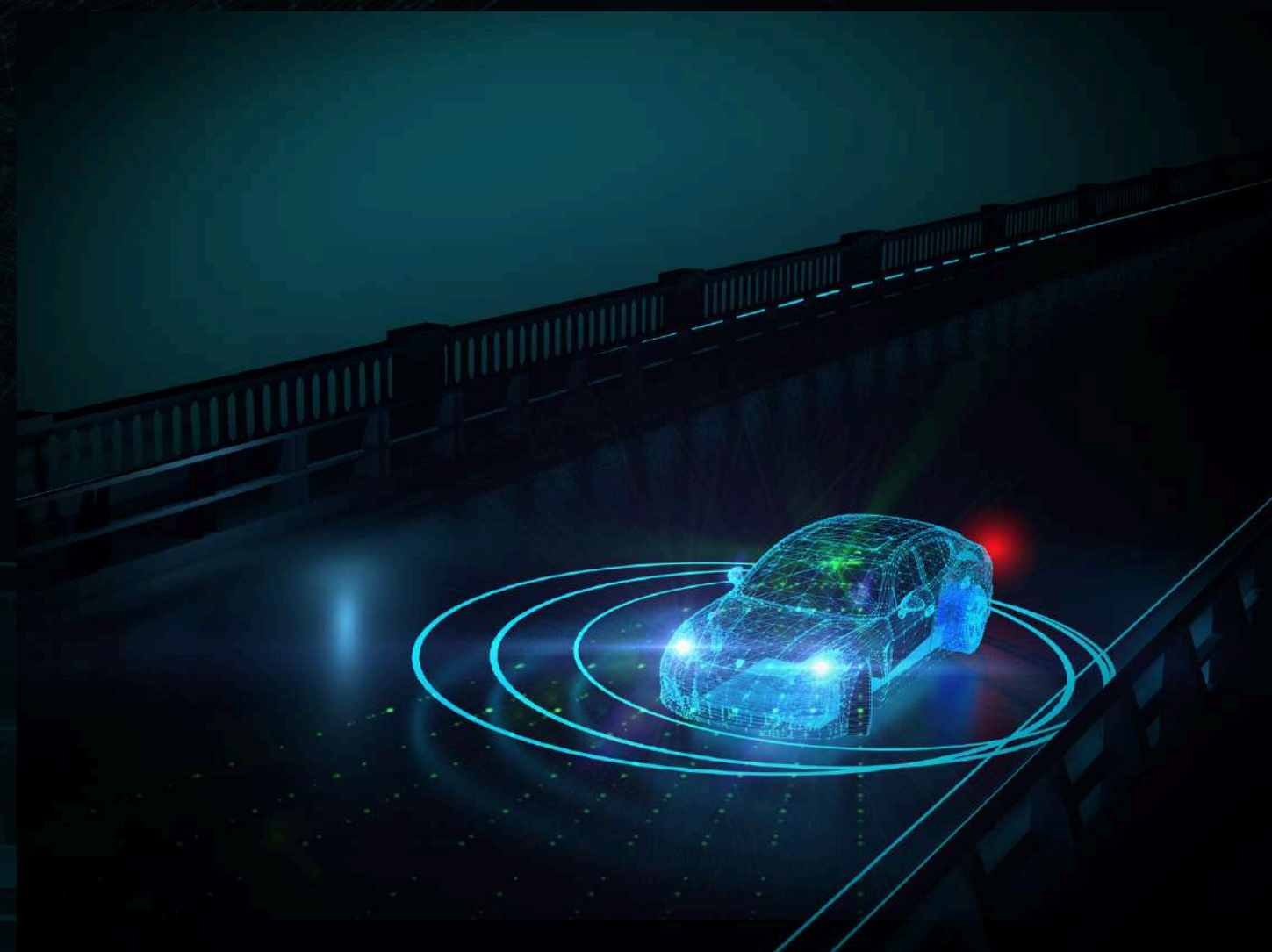
Reconocimiento de Señales de Tráfico

Las CNN también identifican señales como alto, límite de velocidad, ceda el paso o zonas escolares. El sistema reconoce la forma, color y contenido de cada señal en tiempo real, incluso si está parcialmente tapada o desgastada. Con esta información, el auto autónomo adapta su comportamiento a las leyes de tránsito sin depender únicamente de mapas o datos previos.



Conducción End-to-End

En algunos enfoques, las CNN pueden aprender a convertir directamente las imágenes capturadas por las cámaras en acciones de manejo, como girar el volante, acelerar o frenar. Este enfoque, conocido como conducción end-to-end, permite que el vehículo aprenda patrones complejos a partir de millones de ejemplos reales. Aunque aún es experimental, demuestra cómo la IA puede transformar la percepción visual en decisiones de conducción sin pasos intermedios.



Casos reales de la industria



Tesla – cámara + redes neuronales

Tesla utiliza cámaras y redes neuronales para interpretar el entorno sin depender de sensores LiDAR. Su sistema analiza las escenas en tiempo real mediante múltiples modelos CNN que detectan autos, peatones, semáforos y la estructura completa de la carretera. Este enfoque prioriza una visión similar a la humana basada en cámaras.



Waymo – combinación cámara + LiDAR + redes 3D

Waymo combina cámaras, LiDAR y radar junto con redes neuronales 3D para crear una percepción más precisa del entorno. Sus CNN procesan datos de profundidad y nubes de puntos, permitiendo identificar distancias, tamaños y trayectorias de los objetos con alta exactitud. Este enfoque multimodal está orientado a la máxima seguridad.



NVIDIA Drive – plataformas CNN optimizadas a tiempo real

NVIDIA Drive ofrece una plataforma especializada que utiliza CNN optimizadas para ejecutarse en tiempo real dentro del vehículo. Estas redes analizan video, señales de sensores y mapas, permitiendo que los fabricantes desarrollen sus propios sistemas de percepción y conducción basados en IA con alto rendimiento y baja latencia.



TESLA



WAYMO



Retos actuales de la IA en autos autónomos

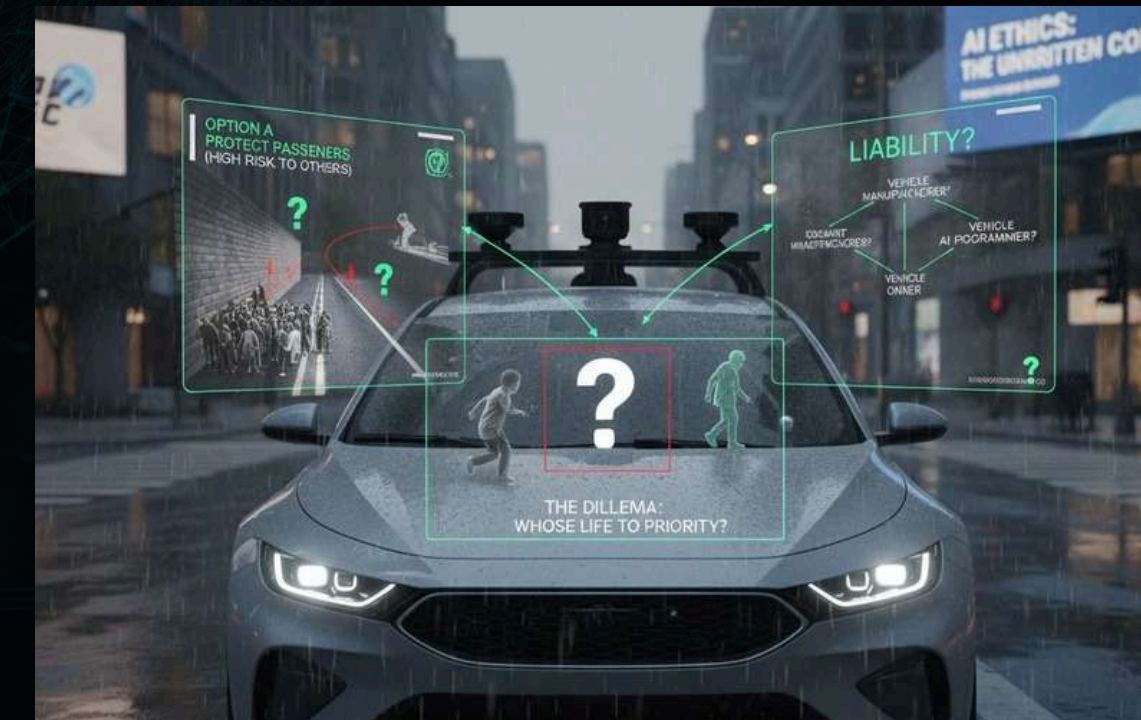


- »» Datos insuficientes para casos raros ("corner cases")



- »» Errores de percepción = accidentes

- »» Condiciones difíciles: lluvia, noche, nieve



- »» Interpretación ética y responsabilidad.