

# VEHÍCULOS AUTÓNOMOS Y PERCEPCIÓN DEL ENTORNO.

## APLICACIONES BASADAS EN DEEP LEARNING.

---

Carlos Guindel



Webminar CEA  
24 de enero de 2019



# Índice

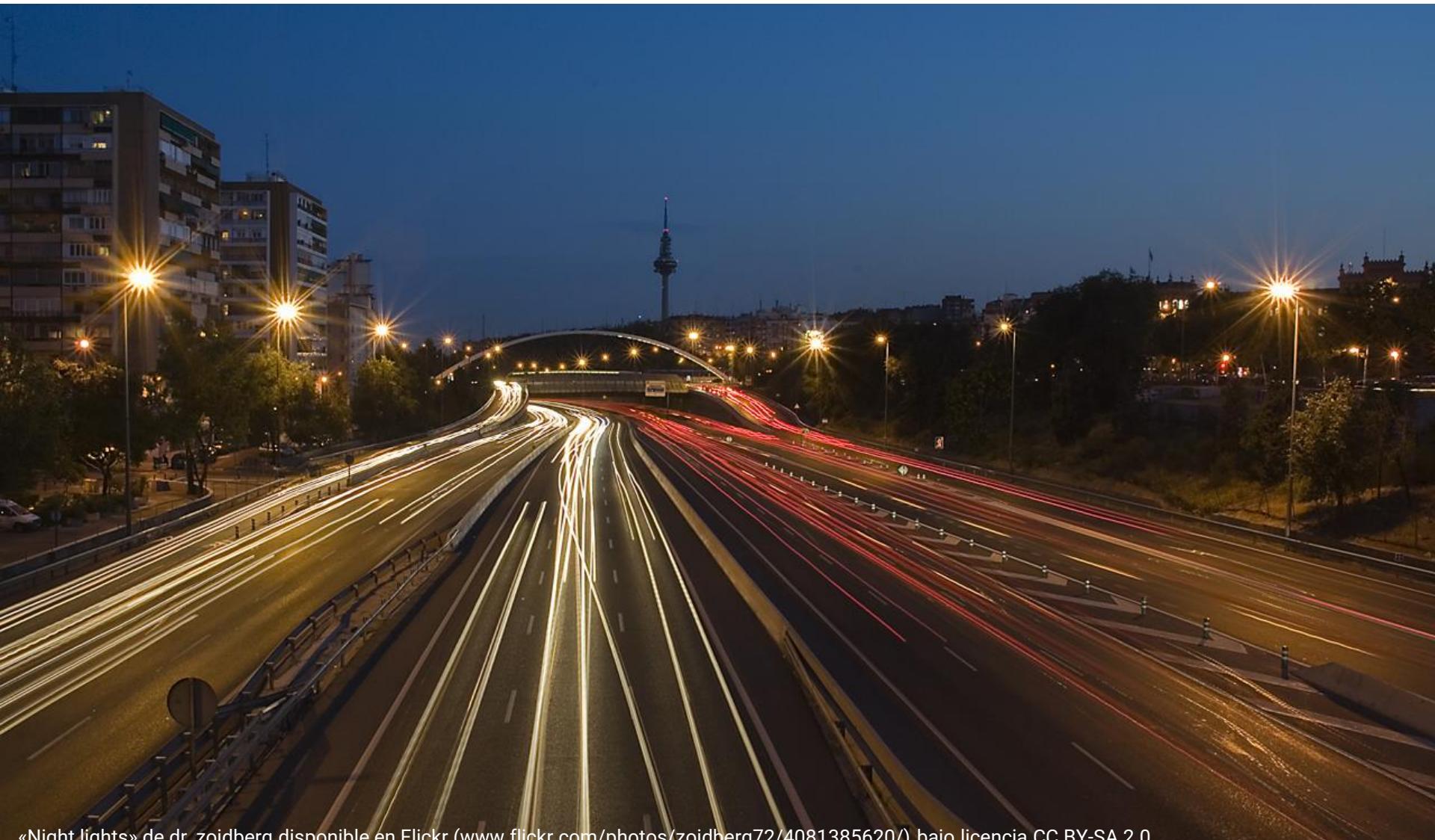
---

- Vehículos autónomos
  - Motivación y antecedentes
  - Taxonomía
- Sistemas de percepción
  - Sensores
  - *Deep learning*
- Aplicaciones
  - Clasificación, detección y localización de objetos
  - Segmentación semántica

---

# Vehículos autónomos

# Transporte



«Night lights» de dr\_zoidberg disponible en Flickr ([www.flickr.com/photos/zoidberg72/4081385620/](http://www.flickr.com/photos/zoidberg72/4081385620/)) bajo licencia CC BY-SA 2.0

# Transporte



# Problemas del transporte

---



## Accidentes

- 1,4 millones de muertes al año en el mundo: más de 9 Boeing 747-400 al día
- Primera causa de muerte en el mundo entre los 5-29 años
- +90% debidos a errores humanos



## Contaminación

- Siete millones de muertes por exposición a partículas en el aire.
- 23% de las emisiones de gases de efecto invernadero.

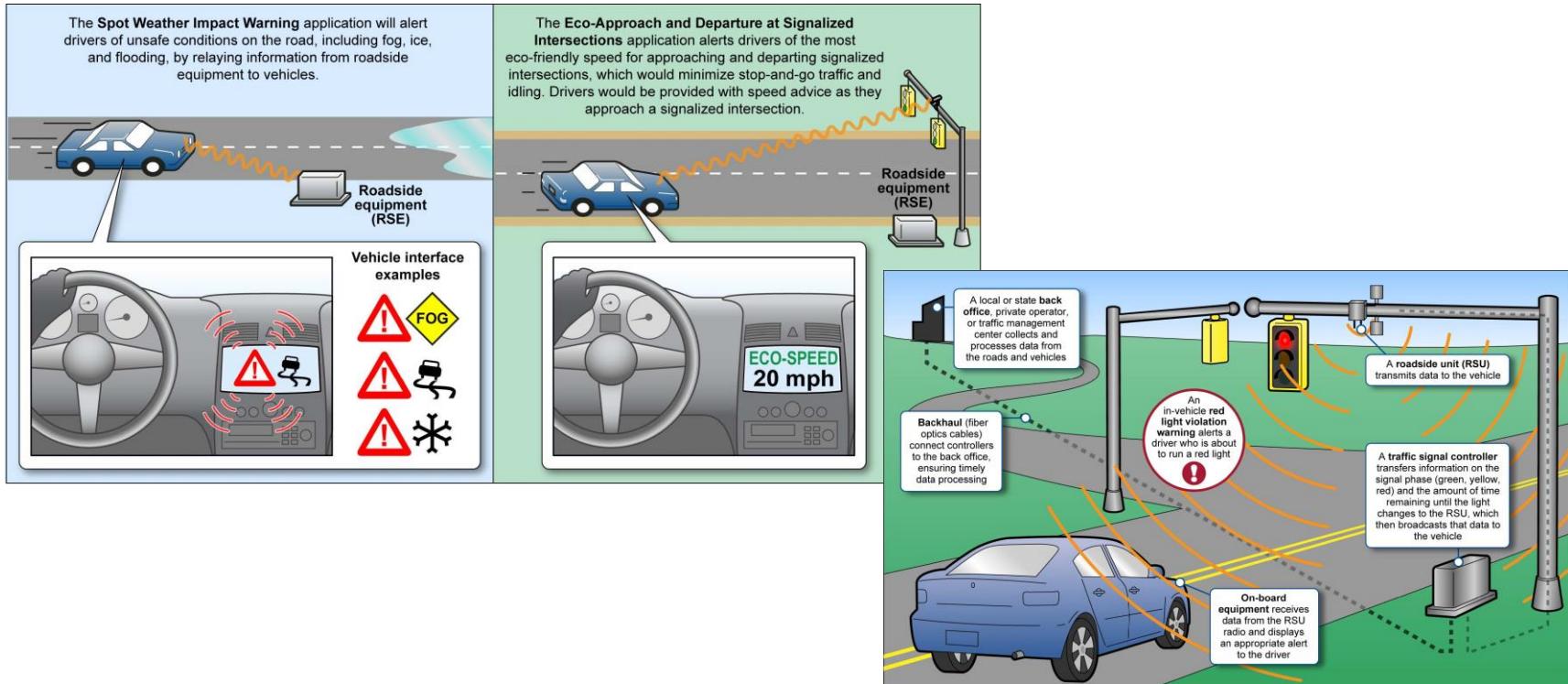


## Atascos

- 2% del PIB en Europa.

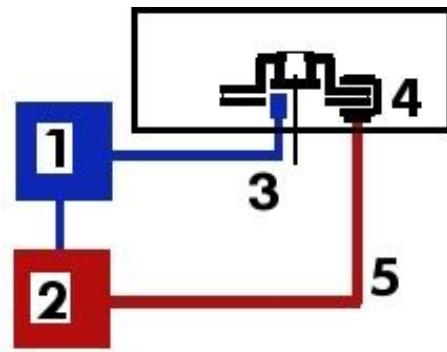
# Sistemas Inteligentes de Transporte (ITS)

- Aplicación de la tecnología para mejorar el transporte (carretera):
  - Seguridad
  - Eficiencia
  - Confort



# Seguridad activa en los vehículos

- La tecnología se ha utilizado desde los inicios del automóvil para reducir y mitigar los accidentes
- Avance empujado por el desarrollo de la electrónica y los sensores



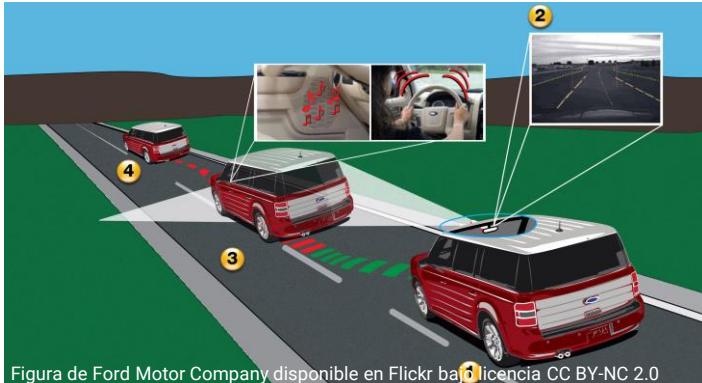
ABS (1978)



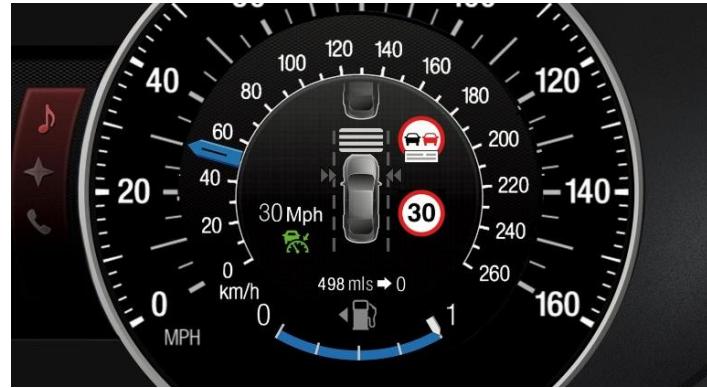
ESP (1990)

# Sistemas de asistencia al conductor

## Sistemas Avanzados de Asistencia al Conductor (ADAS)



Salida de carril (LDW/LKA)



Asistente velocidad (ISA)



Freno de emergencia (AEB)



Detección de punto ciego

# Vehículo autónomo

Un vehículo autónomo es aquel capaz de realizar todas las tareas necesarias para moverse de un punto a otro sin necesidad de intervención humana



# Vehículo autónomo

- Eliminar al conductor de la ecuación
- Potenciales beneficios:



Reducción de accidentes



Uso más eficiente de las infraestructuras



Uso más eficiente del tiempo



Mejora de la productividad



Acceso al transporte sin restricciones



Cambio de paradigma



Nuevos modelos de negocio

# ¿Ciencia ficción?

- Fuertes inversiones de fabricantes:
  - Ford: 1.000 millones \$ en Argo AI (2017)
  - Toyota: 2.800 millones \$ en Toyota Research Institute (2018)
  - etc.
- Predicciones de los expertos:
  - Capricho para ricos alrededor de los 2020-30
  - Adopción generalizada a finales de los 2030-inicio de los 2040

Fuente: T. Litman, "Autonomous vehicle implementation predictions. Implications for Transport Planning," Victoria Transport Policy Institute, 2018.



Foto de Dllu disponible en Wikimedia bajo licencia CC BY-SA 4.0



Foto de Metropolitan Area Council disponible en Flickr bajo licencia CC BY 2.0

# Antecedentes

- Primeros desarrollos en los 1980-90
- **DARPA Challenges** 2004-2007
  - Promovidos por EE UU para automatizar sus vehículos militares
  - Participación de universidades y empresas
  - Desierto (2004, 2005) y entorno urbano (2007)



Foto de Rob NREC disponible en Wikimedia bajo licencia CC BY-SA 3.0

# Antecedentes

---

## Google Self-driving Car (ahora Waymo)

- En 2010, Google desvela sus coches autónomos: han recorrido 140 000 millas sin accidentes.
- 2018: Más de 8 millones de km recorridos en carreteras reales
- Waymo One: servicio de “robotaxis” en Phoenix



Foto de Dllu disponible en Wikimedia bajo licencia CC BY-SA 4.0

## Demostraciones públicas alrededor del mundo

- PROUD (VisLab): en tráfico real en Parma (Italia) **2013**
- Bertha Benz Memorial Route (Daimler y KIT): 103 km **2013**
- BB8 (NVIDIA): 98% de autonomía **2016**

# A día de hoy...

- *Pilotos automáticos:*
  - Tesla Autopilot
  - Mercedes Drive Pilot
  - ...
- Funciones:
  - Mantenimiento de carril
  - Control de crucero adaptativo
  - Frenado de emergencia

**Requieren la atención constante del conductor**



Foto de Marc van der Chijs disponible en Flickr bajo licencia CC BY-ND 2.0



Foto de EurovisionNim disponible en Wikimedia bajo licencia CC BY-SA 4.0

# Niveles

- Definidos por SAE International (desarrolladora de estándares)

Nivel	Nombre	Aceleración, dirección,...	Supervisión del entorno	Respuesta a errores	Modos de conducción
0	Sin automatización	Conductor	Conductor	Conductor	No
1	Asistencia al conductor	Conductor y sistema	Conductor	Conductor	Algunos
2	Automatización parcial	<b>Sistema</b>	Conductor	Conductor	Algunos
3	Automatización condicional	Sistema	<b>Sistema</b>	Conductor	Algunos
4	Automatización alta	Sistema	Sistema	<b>Sistema</b>	Algunos
5	Automatización completa	Sistema	Sistema	Sistema	<b>Todos</b>

# Niveles: situación actual (2019)

- **Nivel 0:** coches sin ADAS

Nivel	Nombre	Aceleración, dirección,...	Supervisión del entorno	Respuesta a errores	Modos de conducción
0	Sin automatización	Conductor	Conductor	Conductor	No
1	Asistencia al conductor	Conductor y sistema	Conductor	Conductor	Algunos
2	Automatización parcial	Sistema	Conductor	Conductor	Algunos
3	Automatización condicional	Sistema	Sistema	Conductor	Algunos
4	Automatización alta	Sistema	Sistema	Sistema	Algunos
5	Automatización completa	Sistema	Sistema	Sistema	Todos

# Niveles: situación actual (2019)

- **Nivel 1:** Algunos ADAS (mantenimiento de carril, crucero adatativo, etc.)

Nivel	Nombre	Aceleración, dirección,...	Supervisión del entorno	Respuesta a errores	Modos de conducción
0	Sin automatización	Conductor	Conductor	Conductor	No
1	Asistencia al conductor	Conductor y sistema	Conductor	Conductor	Algunos
2	Automatización parcial	Sistema	Conductor	Conductor	Algunos
3	Automatización condicional	Sistema	Sistema	Conductor	Algunos
4	Automatización alta	Sistema	Sistema	Sistema	Algunos
5	Automatización completa	Sistema	Sistema	Sistema	Todos

# Niveles: situación actual (2019)

- **Nivel 2:** “Pilotos automáticos” actuales

Nivel	Nombre	Aceleración, dirección,...	Supervisión del entorno	Respuesta a errores	Modos de conducción
0	Sin automatización	Conductor	Conductor	Conductor	No
1	Asistencia al conductor	Conductor y sistema	Conductor	Conductor	Algunos
2	Automatización parcial	<b>Sistema</b>	Conductor	Conductor	Algunos
3	Automatización condicional	Sistema	<b>Sistema</b>	Conductor	Algunos
4	Automatización alta	Sistema	Sistema	<b>Sistema</b>	Algunos
5	Automatización completa	Sistema	Sistema	Sistema	<b>Todos</b>

# Niveles: situación actual (2019)

- **Nivel 3 y siguientes:** Solo prototipos

Nivel	Nombre	Aceleración, dirección,...	Supervisión del entorno	Respuesta a errores	Modos de conducción
0	Sin automatización	Conductor	Conductor	Conductor	No
1	Asistencia al conductor	Conductor y sistema	Conductor	Conductor	Algunos
2	Automatización parcial	Sistema	Conductor	Conductor	Algunos
3	Automatización condicional	Sistema	<b>Sistema</b>	Conductor	Algunos
4	Automatización alta	Sistema	Sistema	<b>Sistema</b>	Algunos
5	Automatización completa	Sistema	Sistema	Sistema	<b>Todos</b>

# Niveles: situación actual (2019)

- ¿Dónde está la brecha?

Nivel	Nombre	Aceleración, dirección,...	Supervisión del entorno	Respuesta a errores	Modos de conducción
0	Sin automatización	Conductor	Conductor	Conductor	No
1	Asistencia al conductor	Conductor y sistema	Conductor	Conductor	Algunos
2	Automatización parcial	<b>Sistema</b>	<b>Conductor</b>	Conductor	Algunos
3	Automatización condicional	Sistema	<b>Sistema</b>	Conductor	Algunos
4	Automatización alta	Sistema	Sistema	<b>Sistema</b>	Algunos
5	Automatización completa	Sistema	Sistema	Sistema	<b>Todos</b>

---

# Sistemas de percepción

# Entendimiento del entorno viario

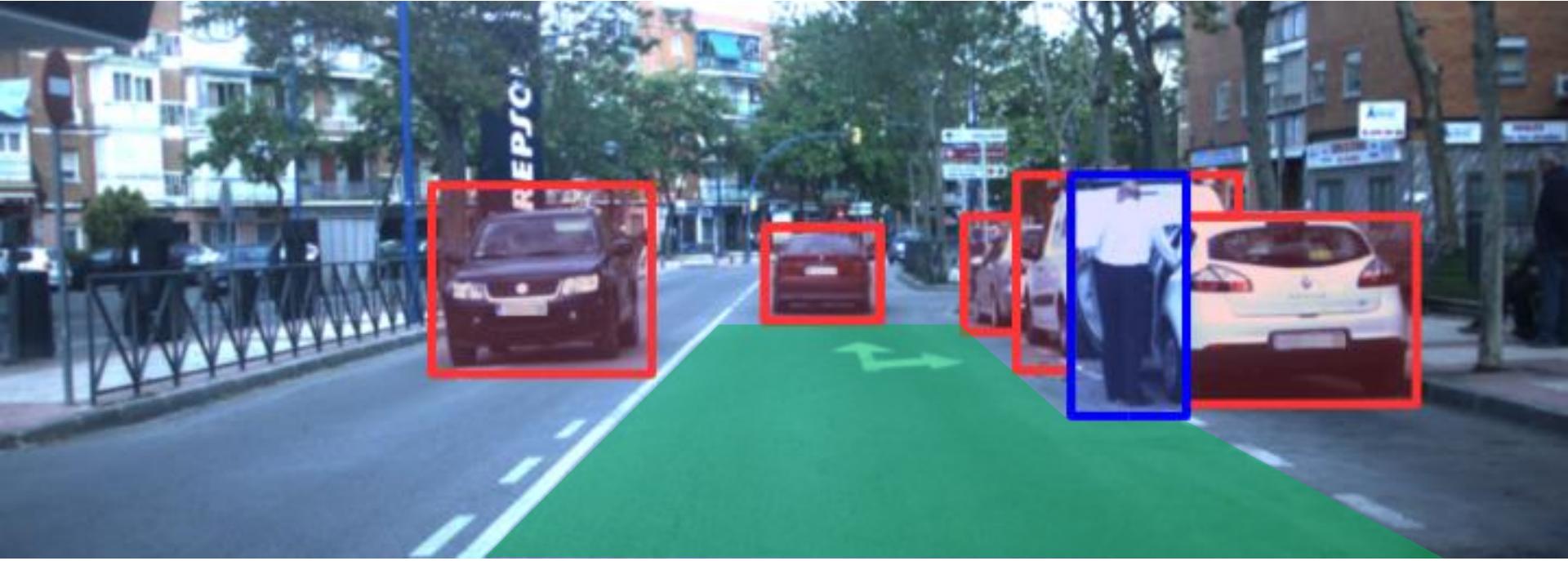
---

- La percepción del entorno es clave, y también la tarea más compleja:



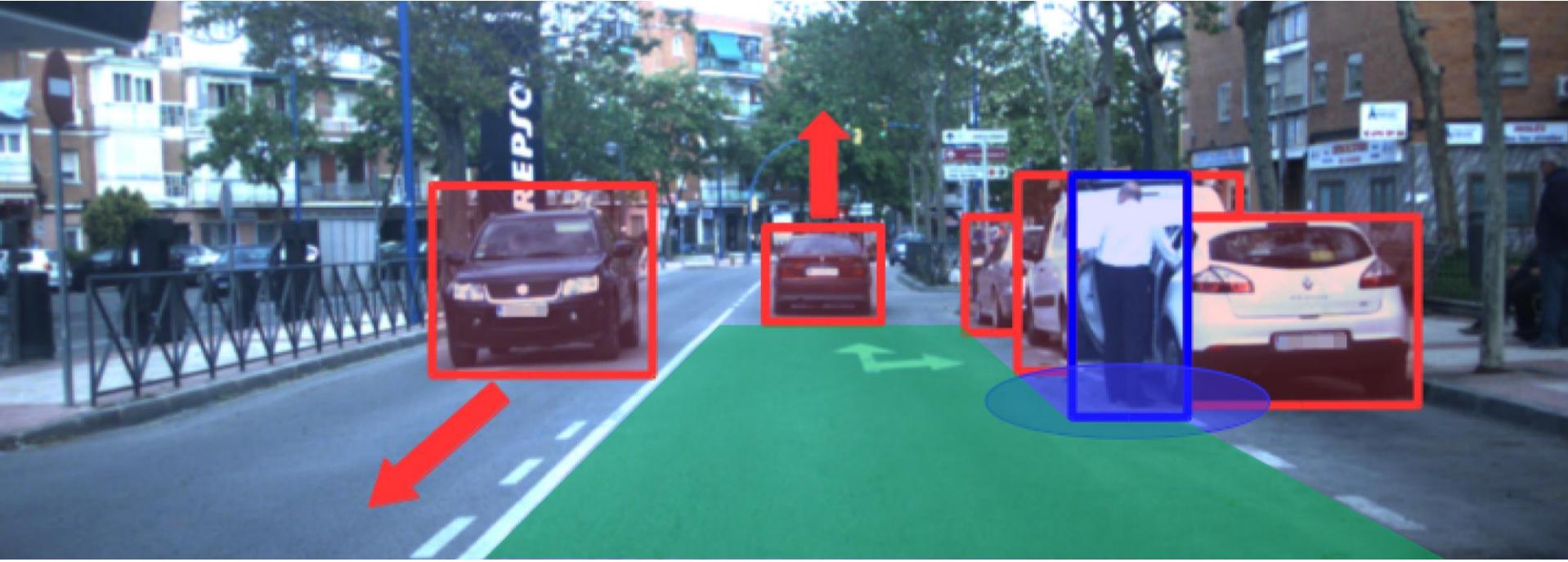
# Entendimiento del entorno viario

- La percepción del entorno es clave, y también la tarea más compleja:



# Entendimiento del entorno viario

- La percepción del entorno es clave, y también la tarea más compleja:



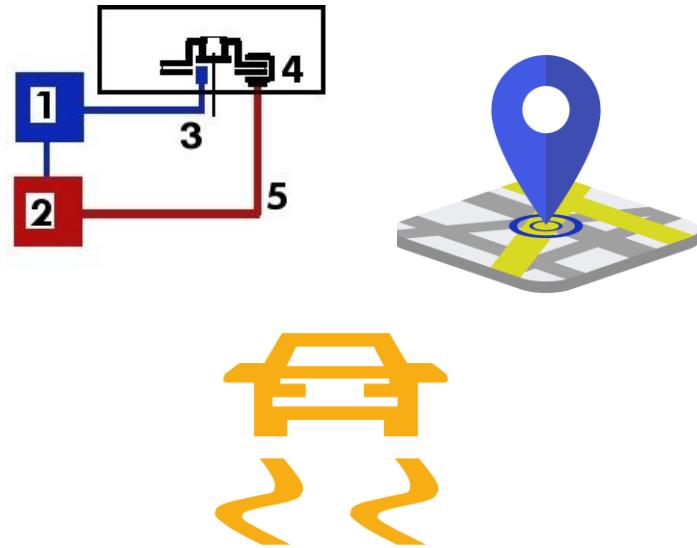
- Variedad de agentes, tamaños y puntos de vista
- Oclusiones
- Condiciones del entorno variables (iluminación, clima, etc.)

Los entornos de tráfico son poco estructurados

# Sensores

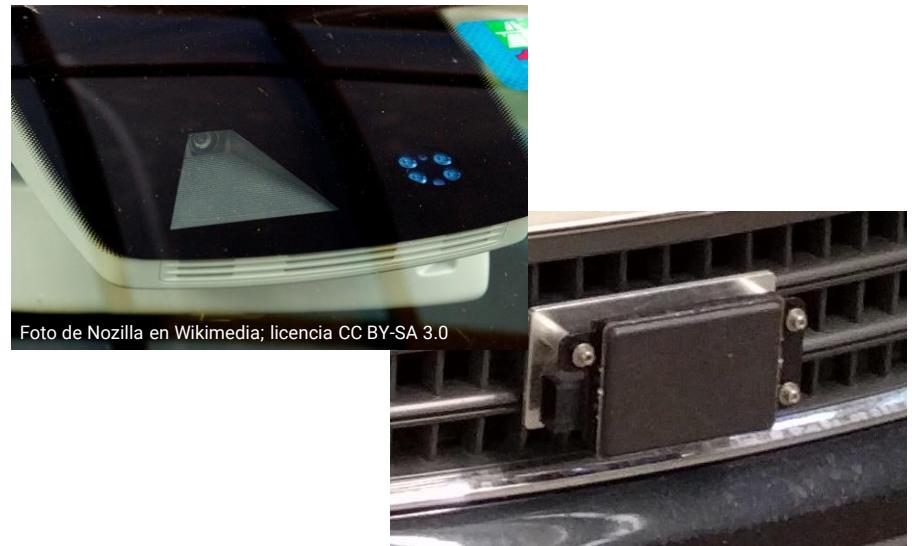
## Propioceptivos

- Miden variables internas
- Ej: deslizamiento de las ruedas, localización del propio vehículo



## Exteroceptivos

- Captan el entorno alrededor del vehículo
- Sistemas de percepción



# Sensores

## Exteroceptivos

### Activos

### Pasivos

- Emiten energía y miden el reflejo
  - Ondas de radio, láser, etc.
- Usan la energía del entorno
  - Luz reflejada en los objetos

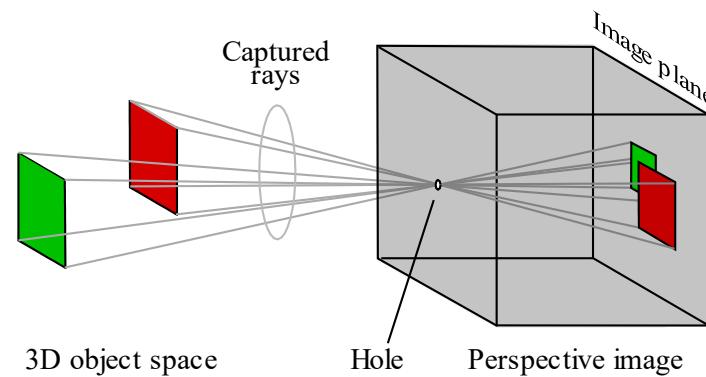
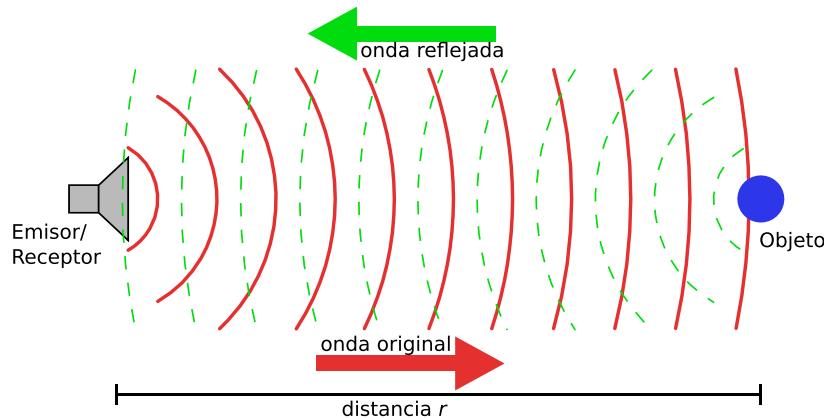


Figura de Georg Wiora en Wikimedia; licencia CC BY-SA 3.0

Ultrasonidos, radar, lidar

Cámaras de vídeo, infrarrojas

# Sensores exteroceptivos activos

## Ultrasonidos

- ✓ Sencillos y baratos
- ✗ Rango muy limitado

Ej.: Asistente aparcamiento

## Radar

- ✓ Mayor rango de detección
- ✗ Información escasa

Ej: ACC, Frenado automático

## Lidar

- ✓ Fiable y robusto
- ✓ Gran rango (hasta 200 m)
- ✗ Elevado coste
- ✗ Información limitada



Foto de Dllu disponible en Wikimedia bajo licencia CC BY-SA 4.0

Proporcionan información de distancias

# Sensores exteroceptivos activos

## Ultrasonidos

- ✓ Sencillos y baratos
- ✗ Rango muy limitado

Ej.: Asistente aparcamiento

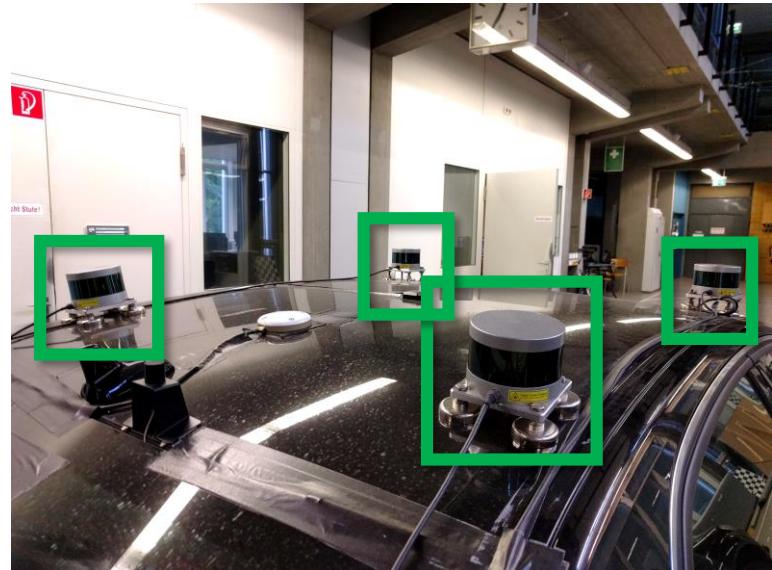
## Radar

- ✓ Mayor rango de detección
- ✗ Información escasa

Ej: ACC, frenado automático

## Lidar

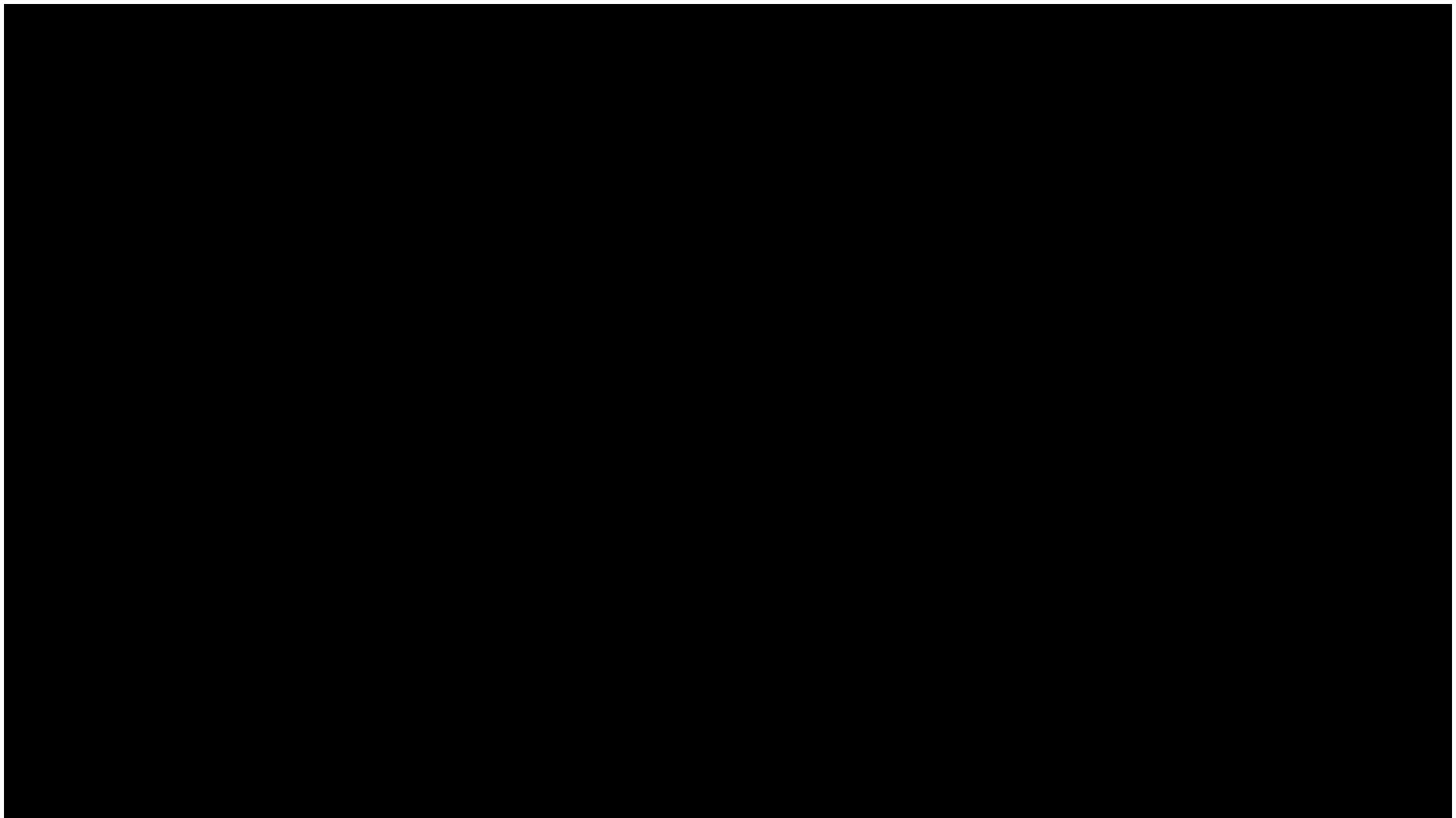
- ✓ Fiable y robusto
- ✓ Gran rango (hasta 200 m)
- ✗ Elevado coste
- ✗ Información limitada



Proporcionan información de distancias

# Lidar

---



# Sensores exteroceptivos pasivos

## Cámaras de vídeo

- ✓ Información rica
- ✓ Poco robusto (iluminación, clima)
- ✗ Dificultad de procesamiento



Foto de Michael Sheenan disponible en Wikimedia bajo licencia CC BY 2.0  
© 2013 HighTechDad

Proporcionan información de apariencia

## Sistemas estéreo

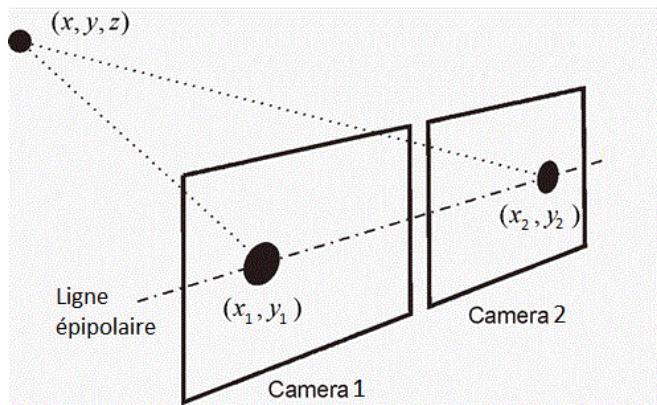


Foto de A. Bailey disponible en Wikimedia bajo licencia CC BY-SA 3.0

# Sensores exteroceptivos pasivos

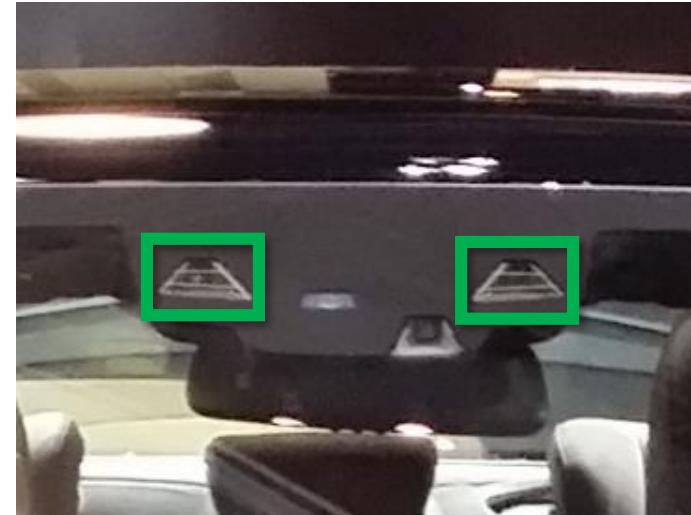
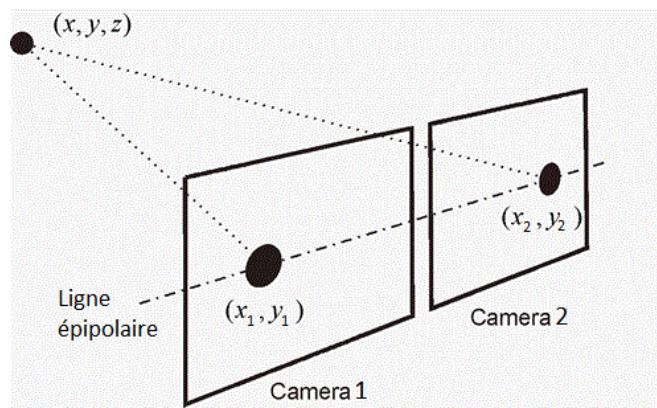
## Cámaras de vídeo

- ✓ Información rica
- ✓ Poco robusto (iluminación, clima)
- ✗ Dificultad de procesamiento



Proporcionan información de apariencia

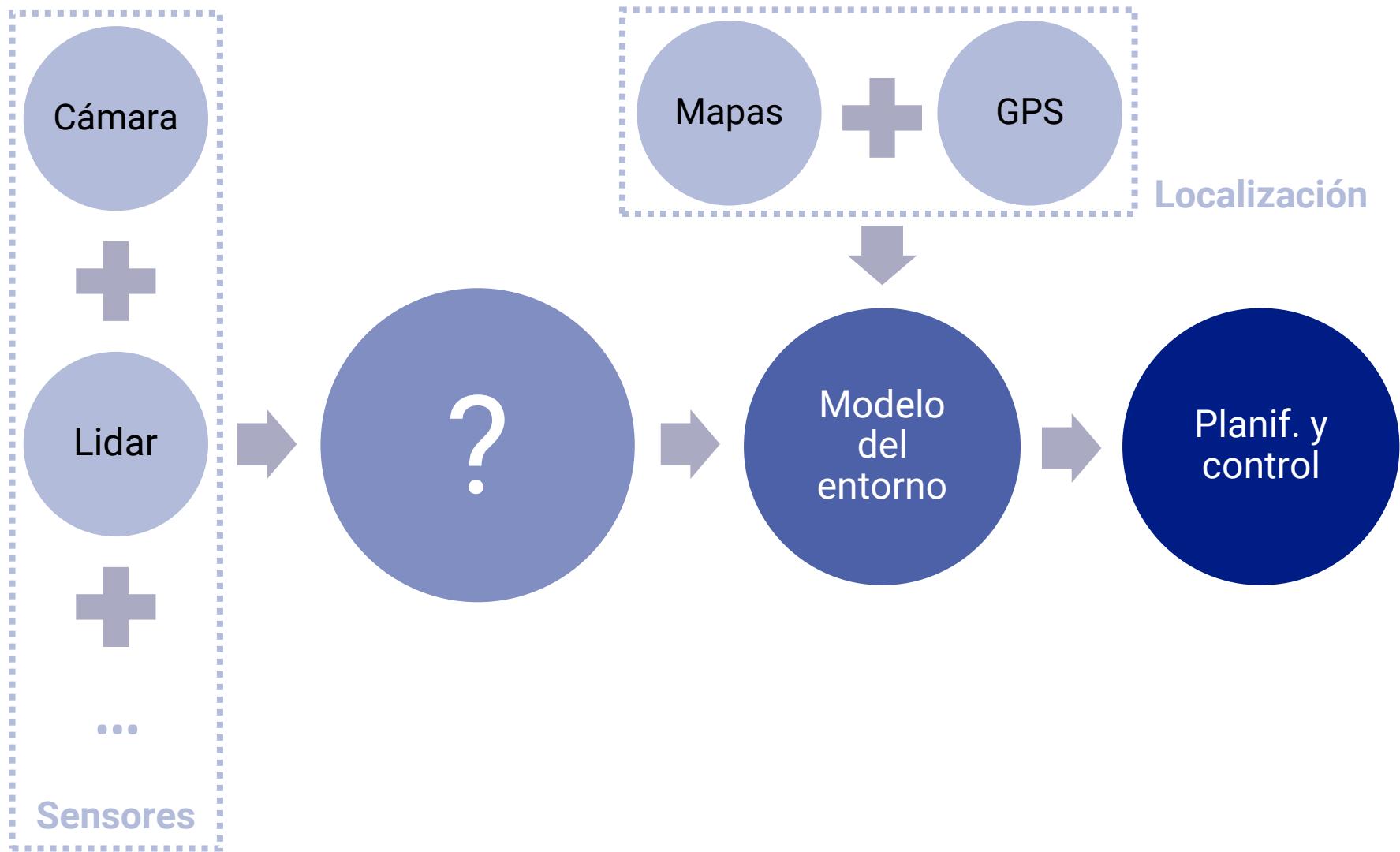
## Sistemas estéreo



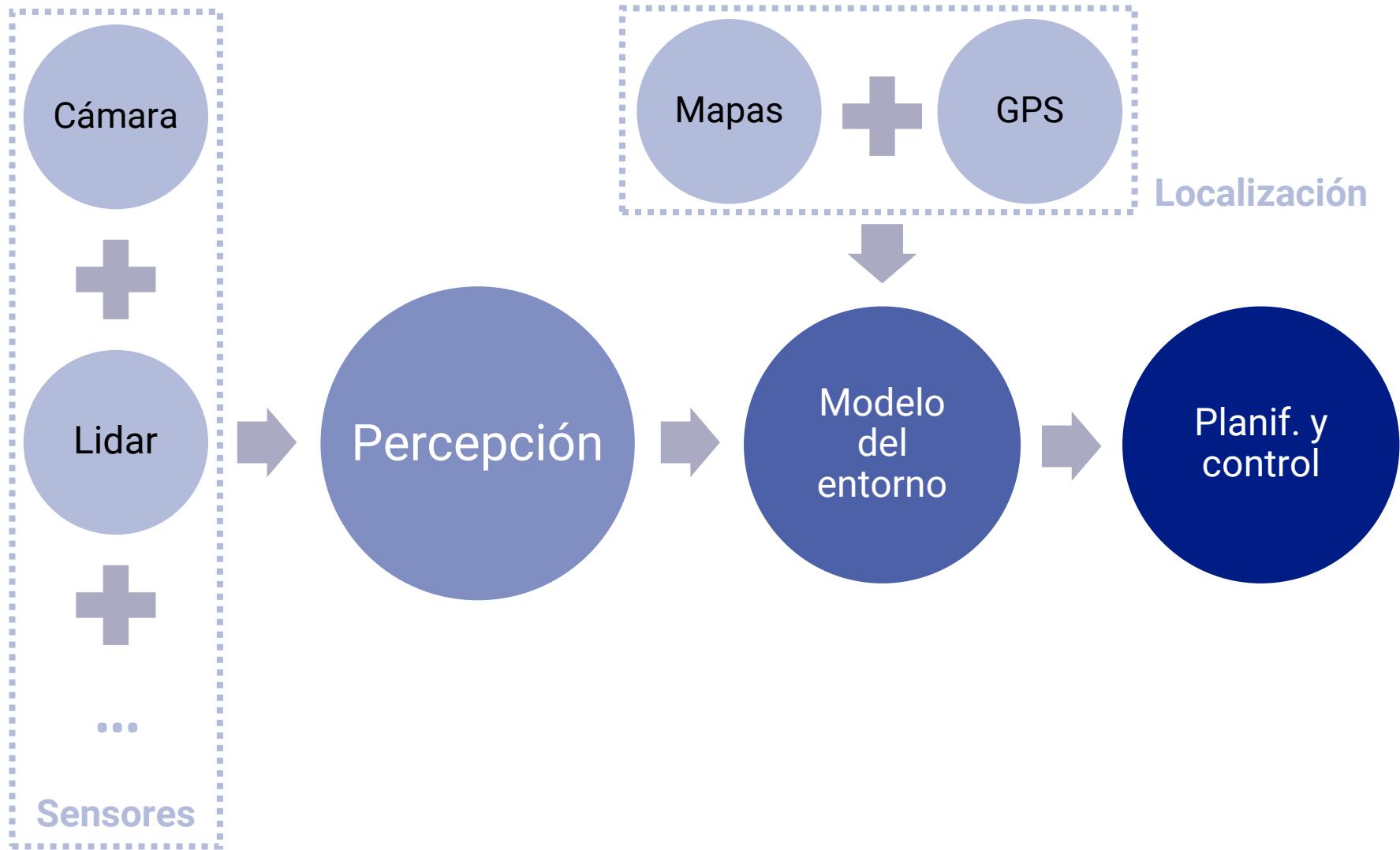
# Visión estéreo



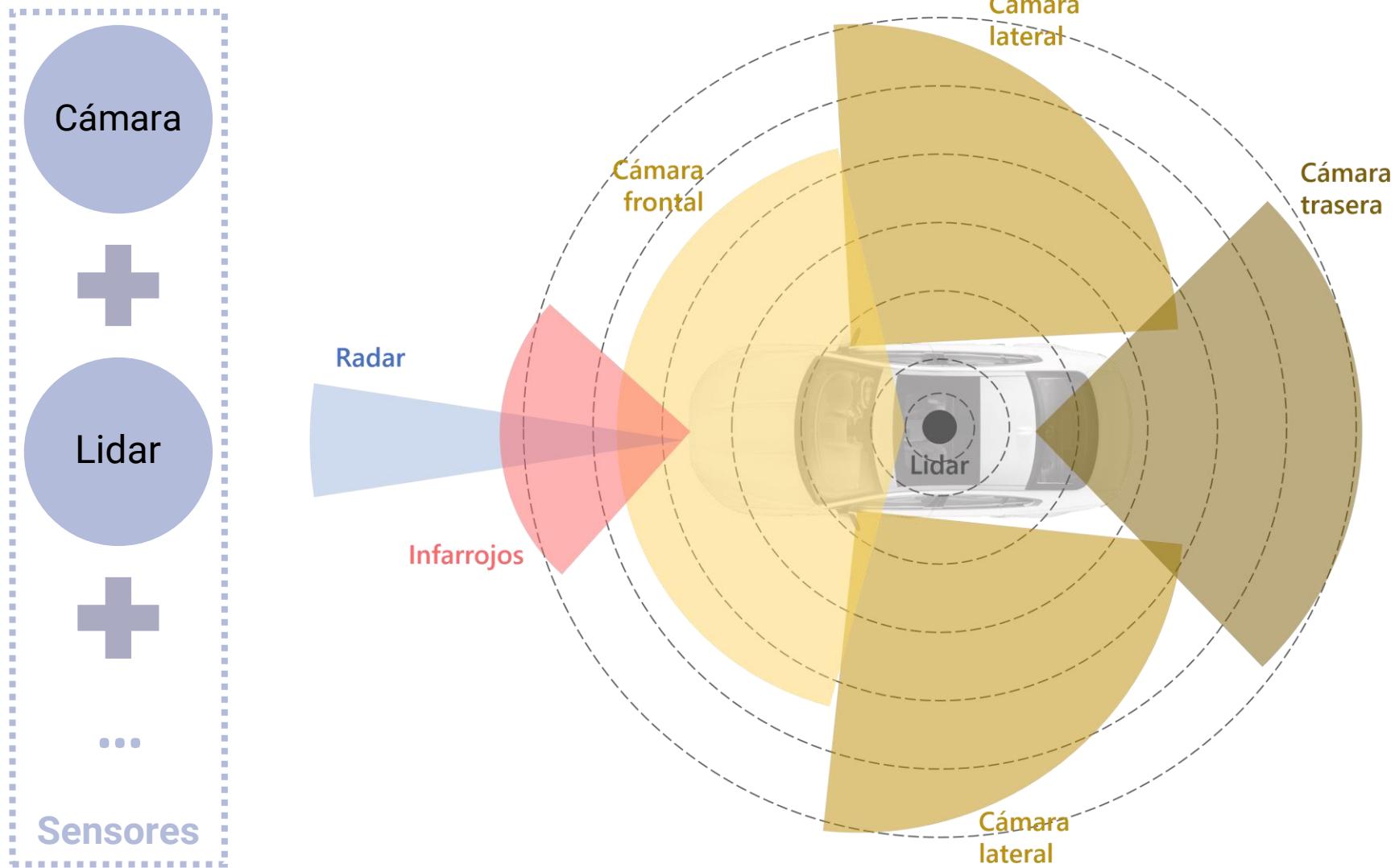
# Conducción autónoma



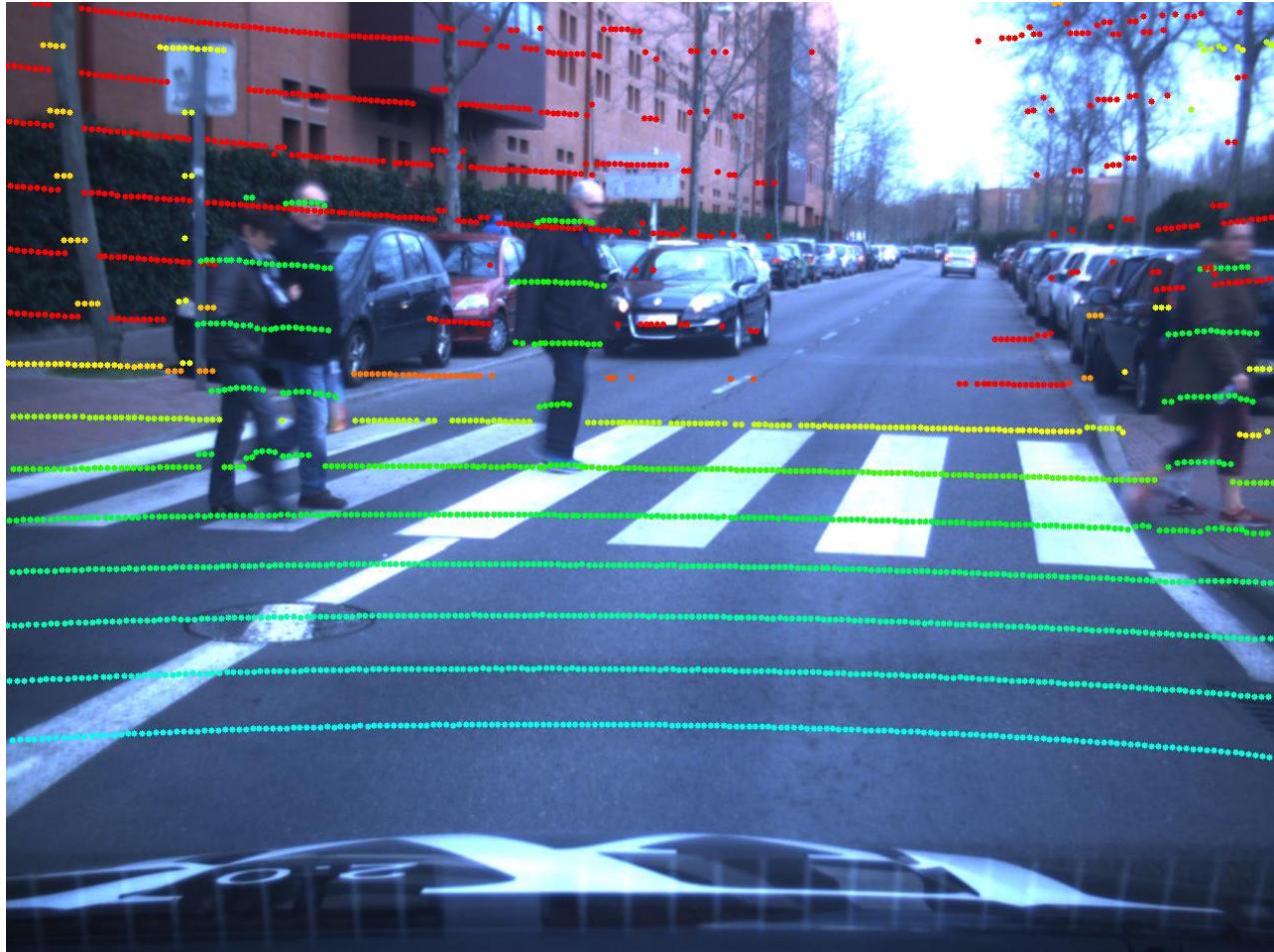
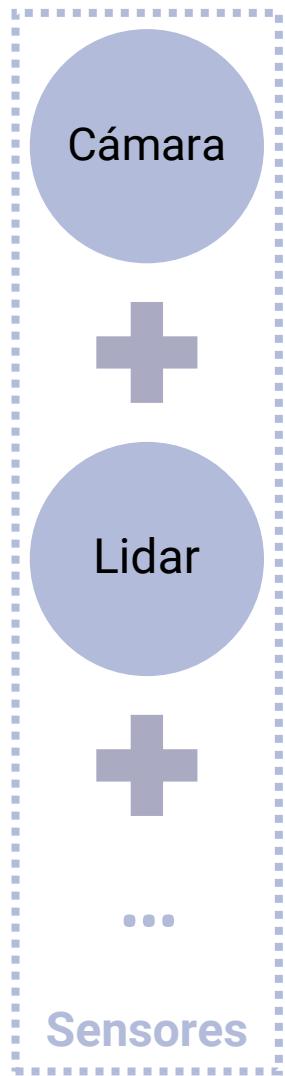
# Conducción autónoma



# Fusión de datos

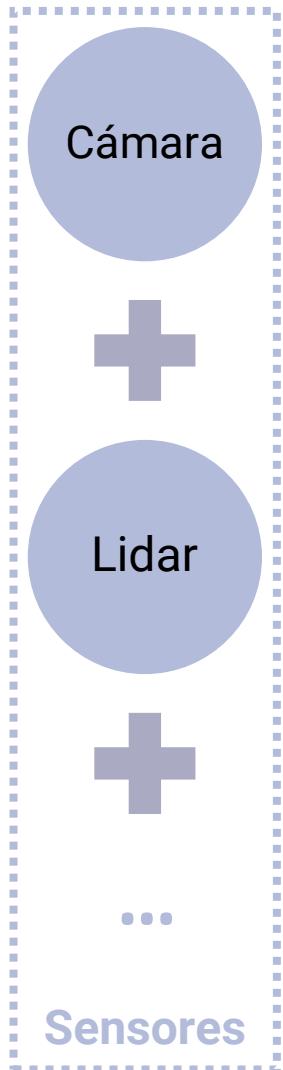


# Fusión de datos y calibración

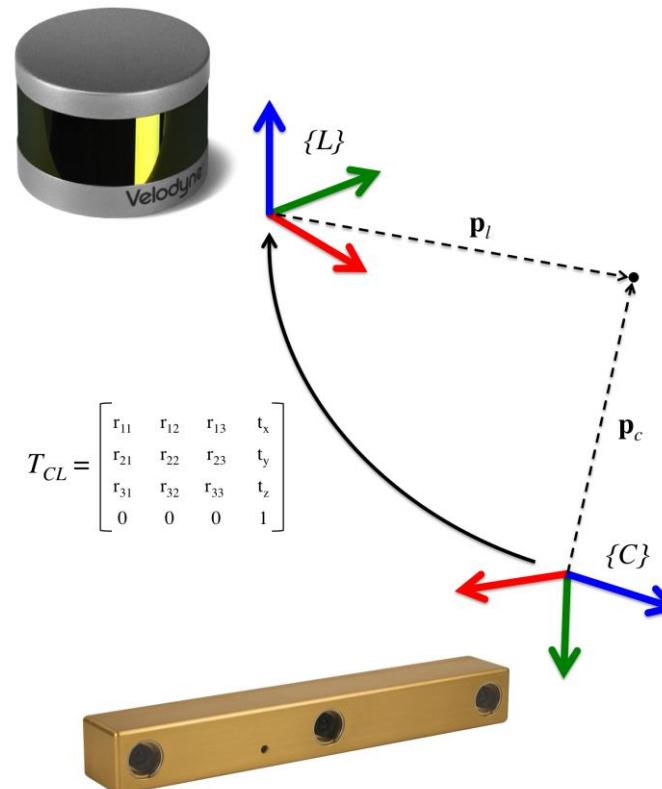


C. Guindel, J. Beltrán, D. Martín, and F. García, "Automatic Extrinsic Calibration for Lidar-Stereo Vehicle Sensor Setups," in IEEE International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 674–679.

# Fusión de datos y calibración



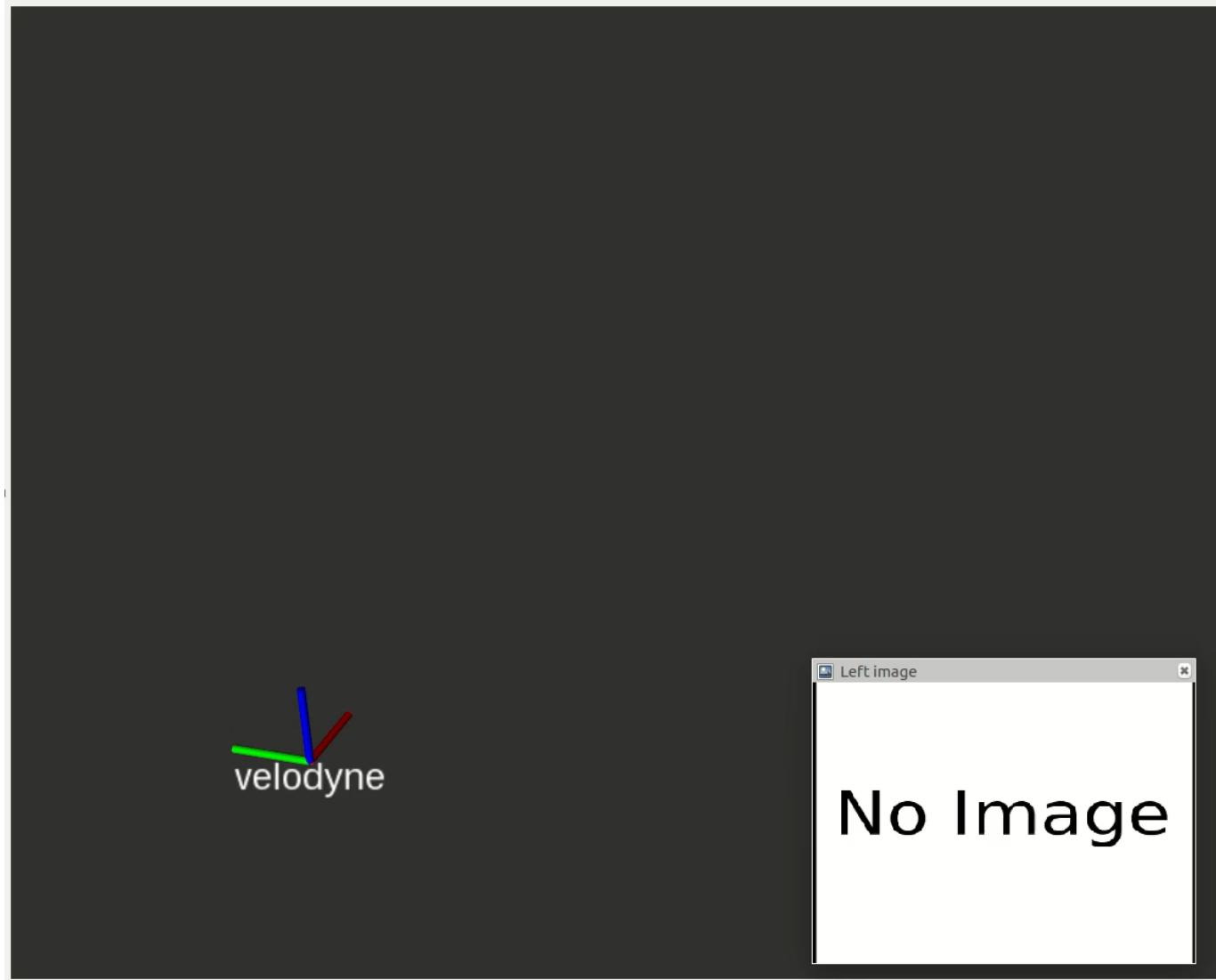
## Parámetros extrínsecos



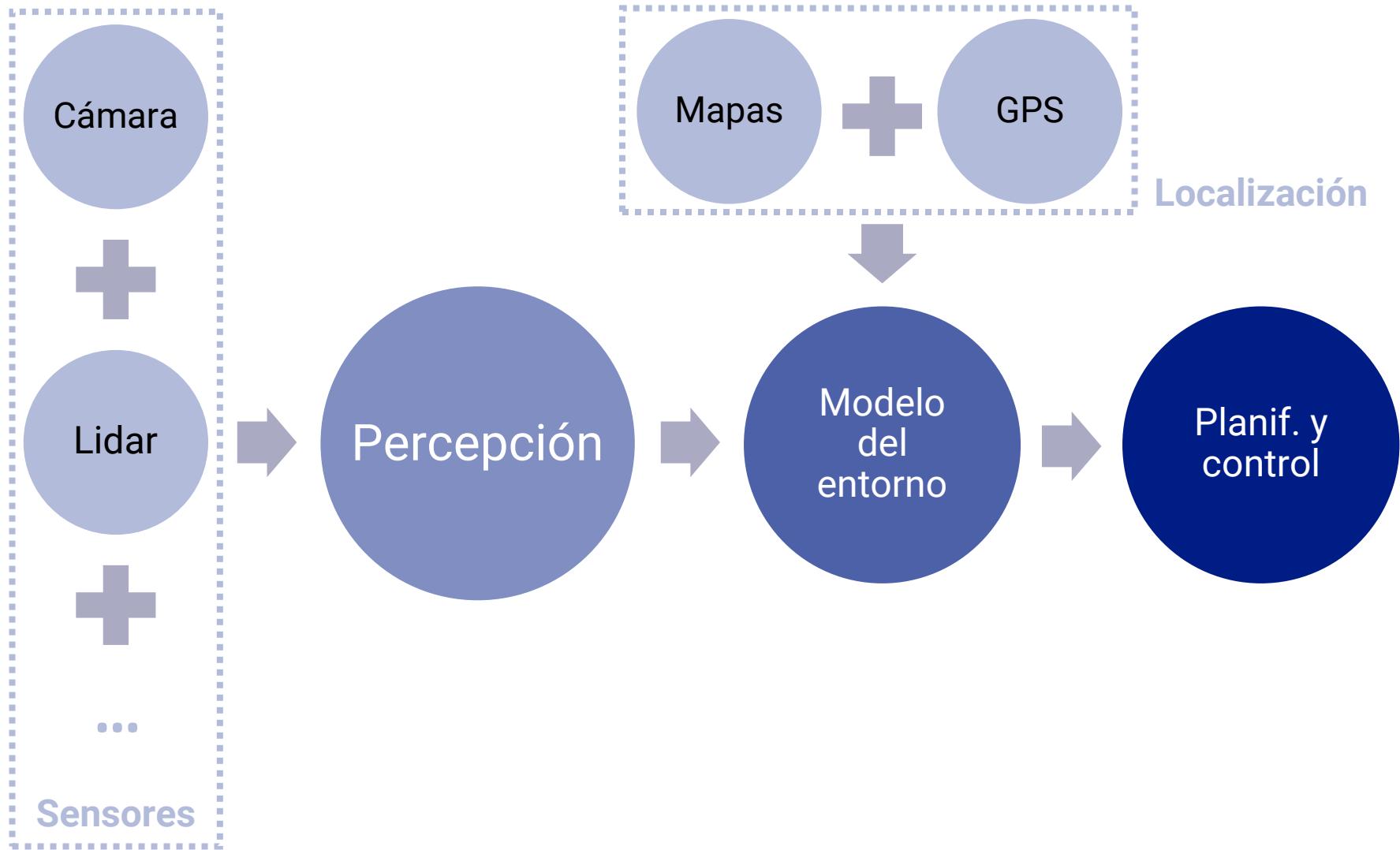
C. Guindel, J. Beltrán, D. Martín, and F. García, "Automatic Extrinsic Calibration for Lidar-Stereo Vehicle Sensor Setups," in IEEE International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 674–679.

# Calibración estéreo-lidar

---

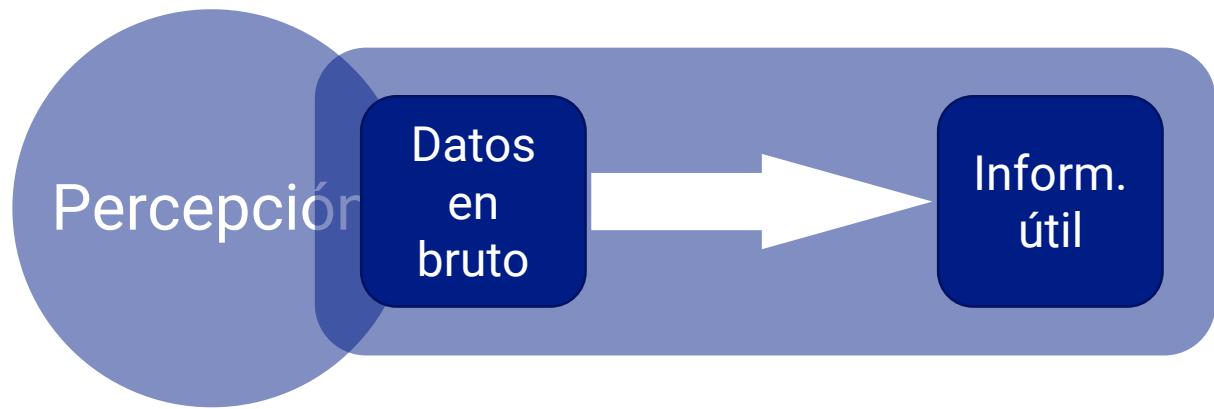


# Conducción autónoma

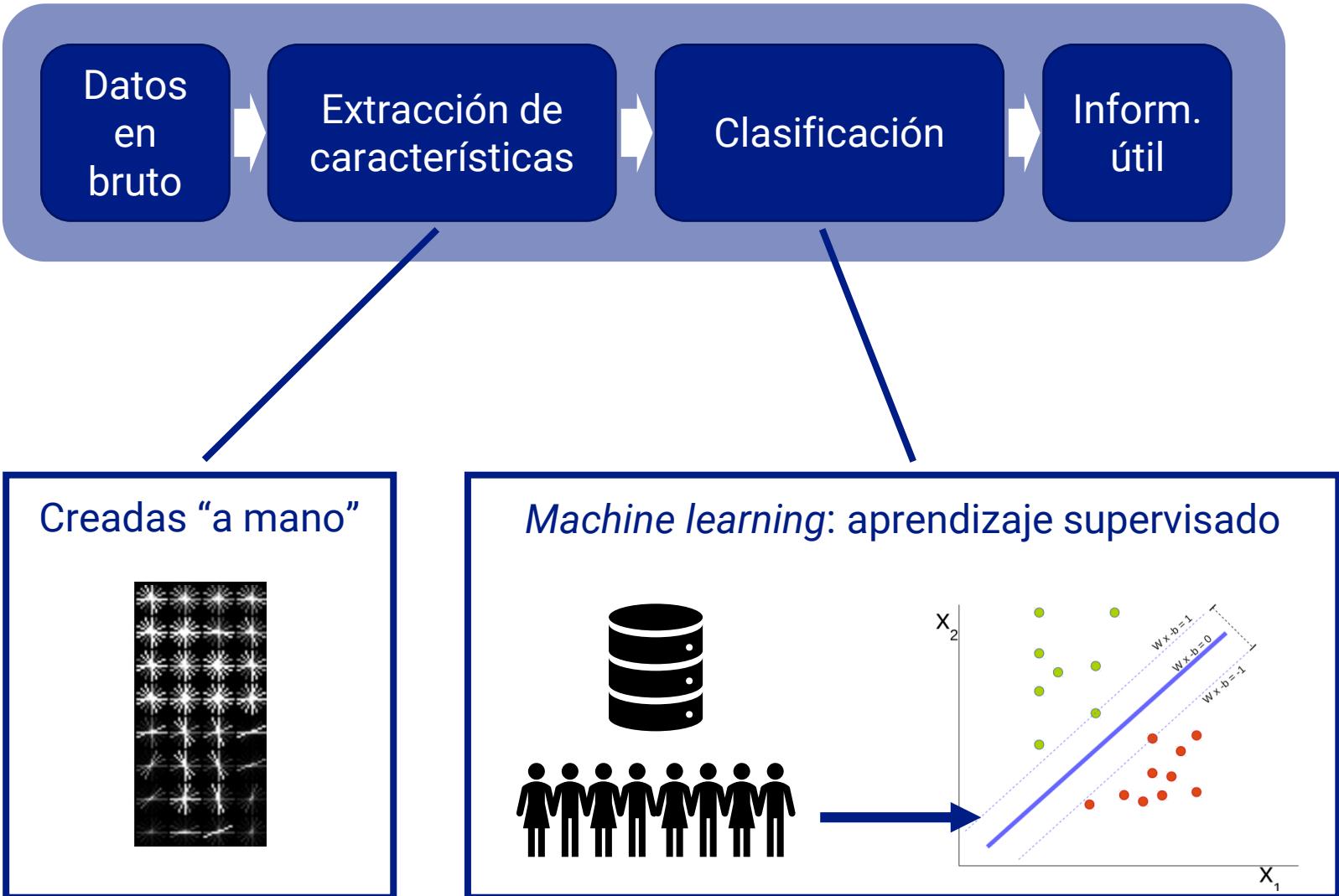


# Conducción autónoma

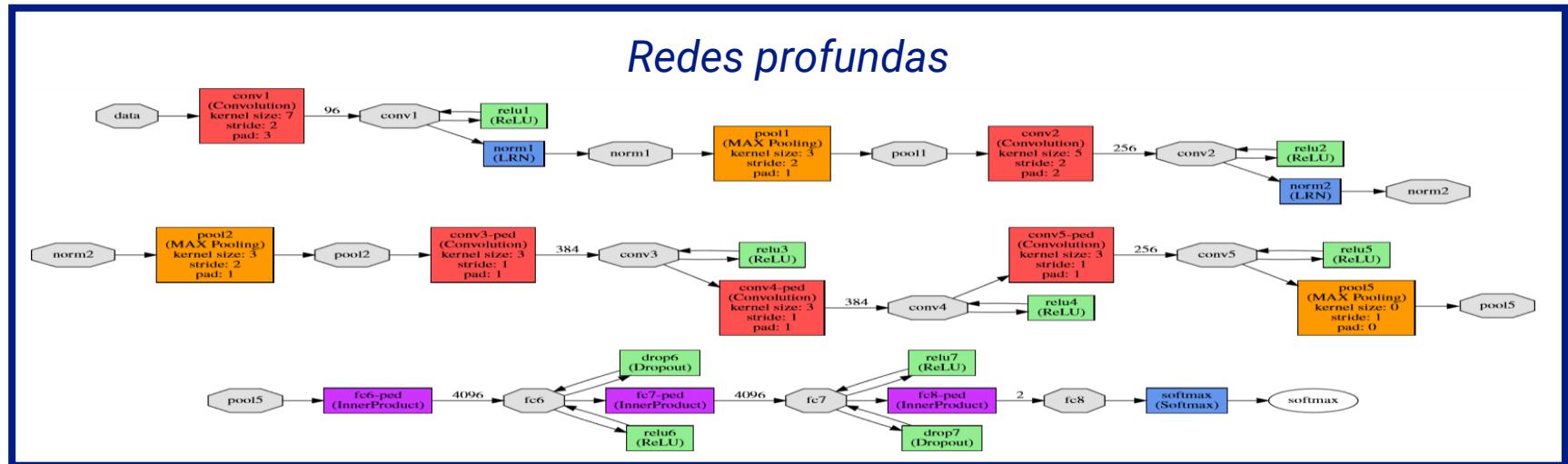
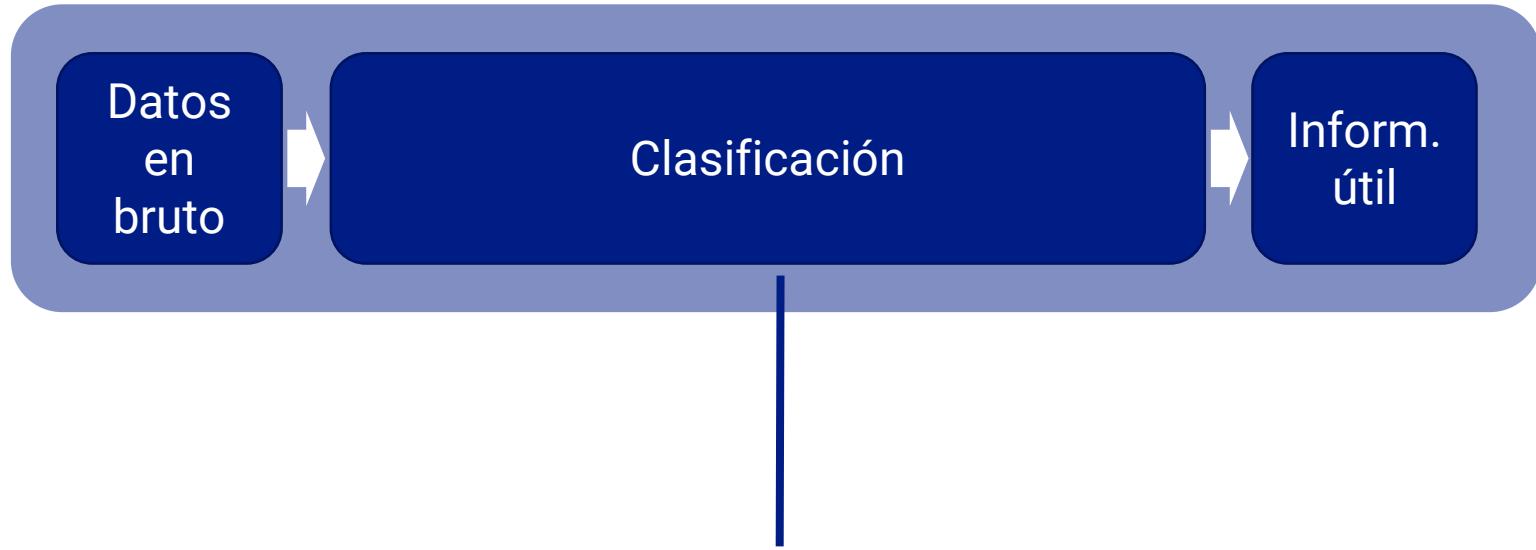
---



# Enfoque clásico

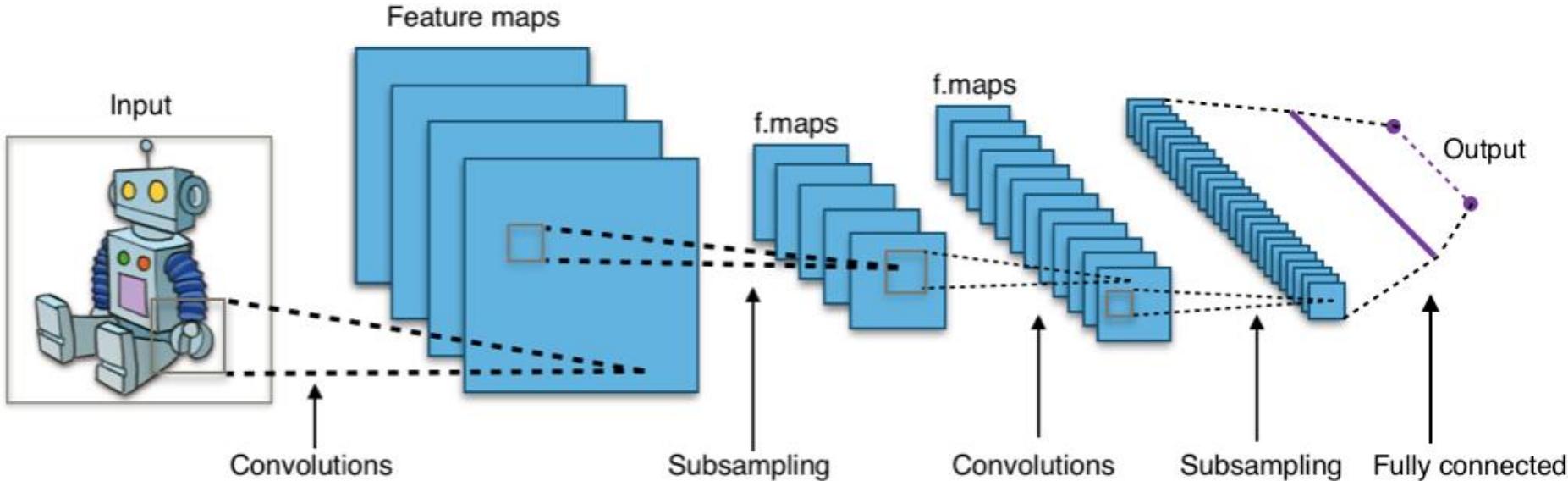


# Aprendizaje de características: *deep learning*



# Redes neuronales convolucionales (CNN)

Para datos con estructura bidimensional (p.ej. imágenes)

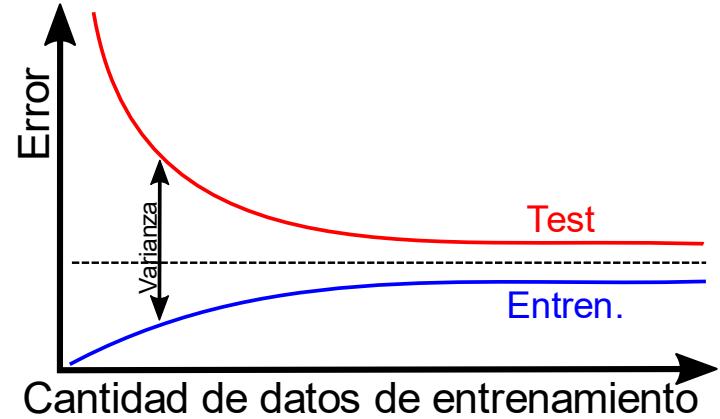
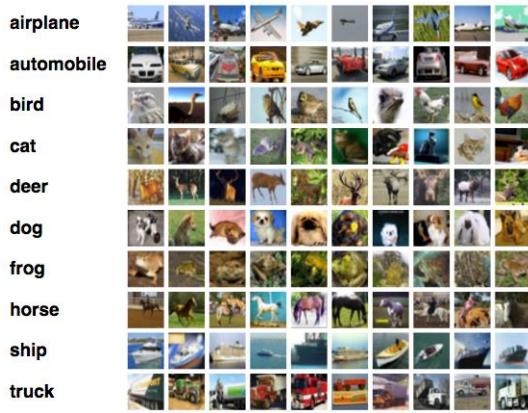


«Typical CNN» de Aphex, disponible en Wikimedia ([https://commons.wikimedia.org/wiki/File:Typical\\_cnn.png](https://commons.wikimedia.org/wiki/File:Typical_cnn.png)) bajo licencia CC BY-SA 4.0

- Sucesión de **convoluciones** con filtros aprendidos, **no linealidades** y **reducciones de muestreo**
- La etapa final está compuesta por redes neuronales clásicas (totalmente conectadas)

# Aprendizaje

- Minimización de una fcn. de coste (error) en el entrenamiento
- Son necesarias muchas muestras etiquetadas



- Bases de datos públicas:
  - Generalistas: ImageNet, COCO,...
  - Conducción: KITTI, Cityscapes, BDD 100k, Apolloscape,...
  - Permiten el entrenamiento y la comparación de algoritmos

**IMAGENET**

[www.image-net.org](http://www.image-net.org)



[cocodataset.org](http://cocodataset.org)

**The KITTI Vision  
Benchmark Suite**

[www.cvlabs.net/datasets/kitti/](http://www.cvlabs.net/datasets/kitti/)



**CITYSCAPES**  
DATASET

[www.cityscapes-dataset.com](http://www.cityscapes-dataset.com)

# Aplicaciones

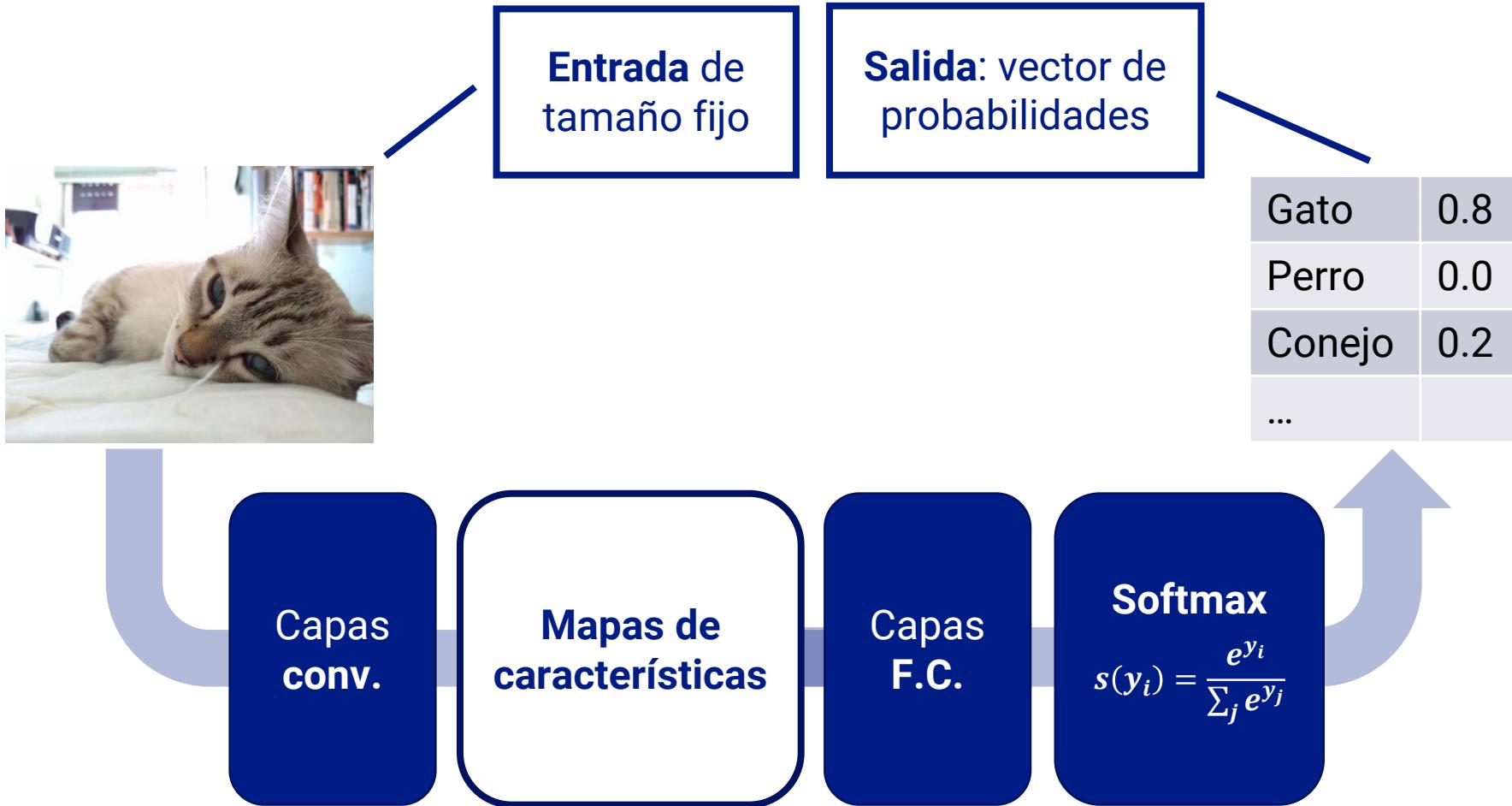


Lab. Sistemas  
Inteligentes



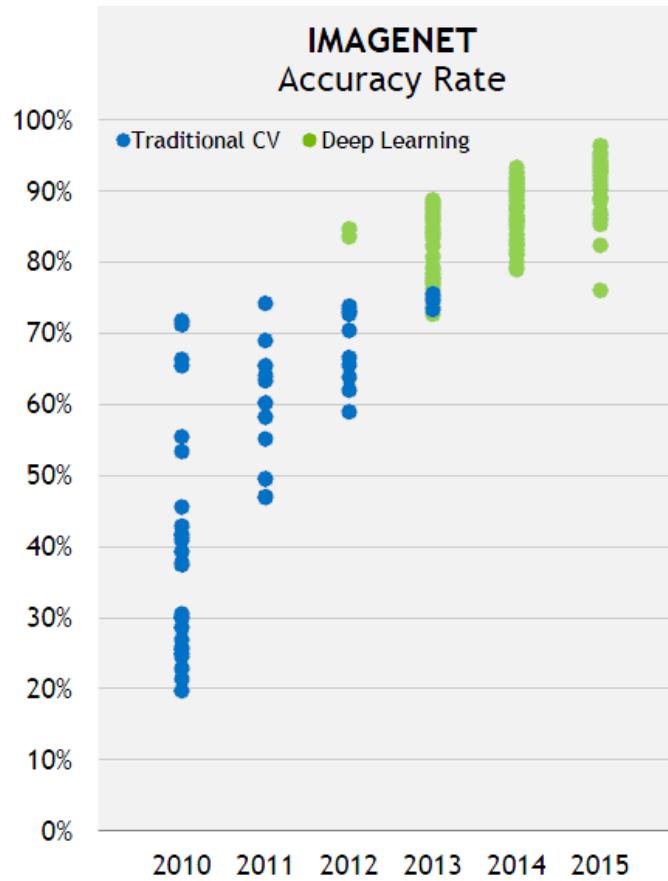
# Clasificación

- Aplicación directa de las CNN



# Clasificación

- Aplicación directa de las CNN



Fuente: NVIDIA CES 2016 Press Conference  
<https://www.slideshare.net/NVIDIA/nvidia-ces-2016-press-conference>

# Clasificación: señales de tráfico

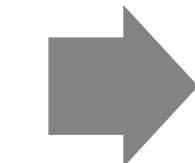


- Basada en el color
- Genera propuestas susceptibles de ser señal

- Aplicación de una CNN por cada propuesta
- Clase “fondo” para errores de segmentación



Regiones



CNN



CNN



“Fondo”

Precisión en la clasificación: **99,05%**

L. Díaz, "Sistema de detección y clasificación de señales de tráfico basado en Deep Learning," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2017

# Clasificación: estado del semáforo



- Mapas digitales
- Posicionamiento usando visión

Regiones precisas



(a) Bosch



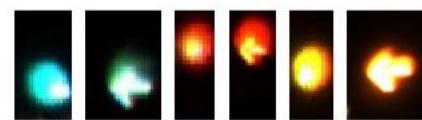
(b) Cityscapes



(c) LaRA



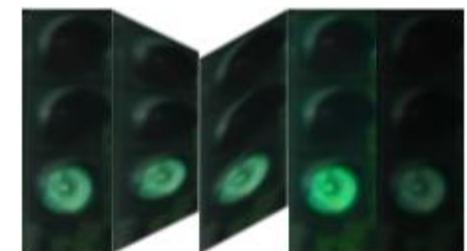
(d) LISA-Day



(e) LISA-Night



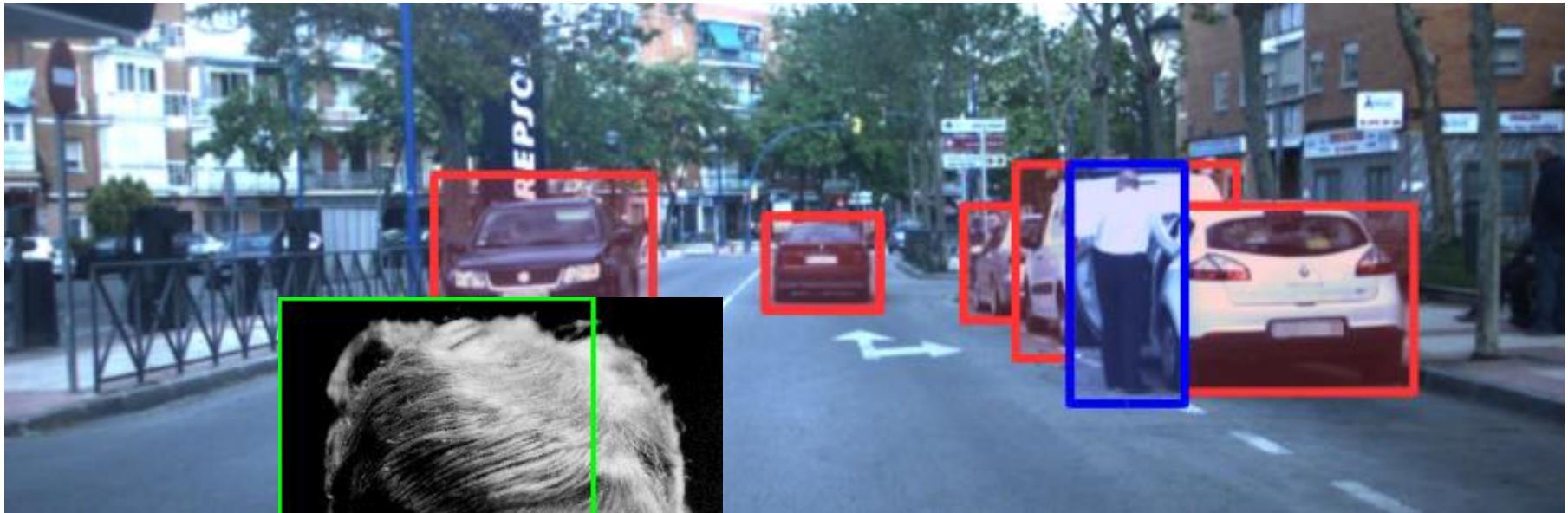
(f) WPI



- Combinación de varias bases de datos + aumentado de datos
- CNN ligera (7 ms)
- F1-score: **90,3%** (vs. 67,6% Random Trees)

C. Fernández, C. Guindel, N.-O. Salscheider, and C. Stiller, "A Deep Analysis of the Existing Datasets for Traffic Light State Recognition," in Proc. IEEE International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 248–254

# Detección



- En las imágenes se detectan los elementos
- Salida de la red: la ubicación del objeto y su posición en la imagen



les elementos

gente y su posición en la imagen

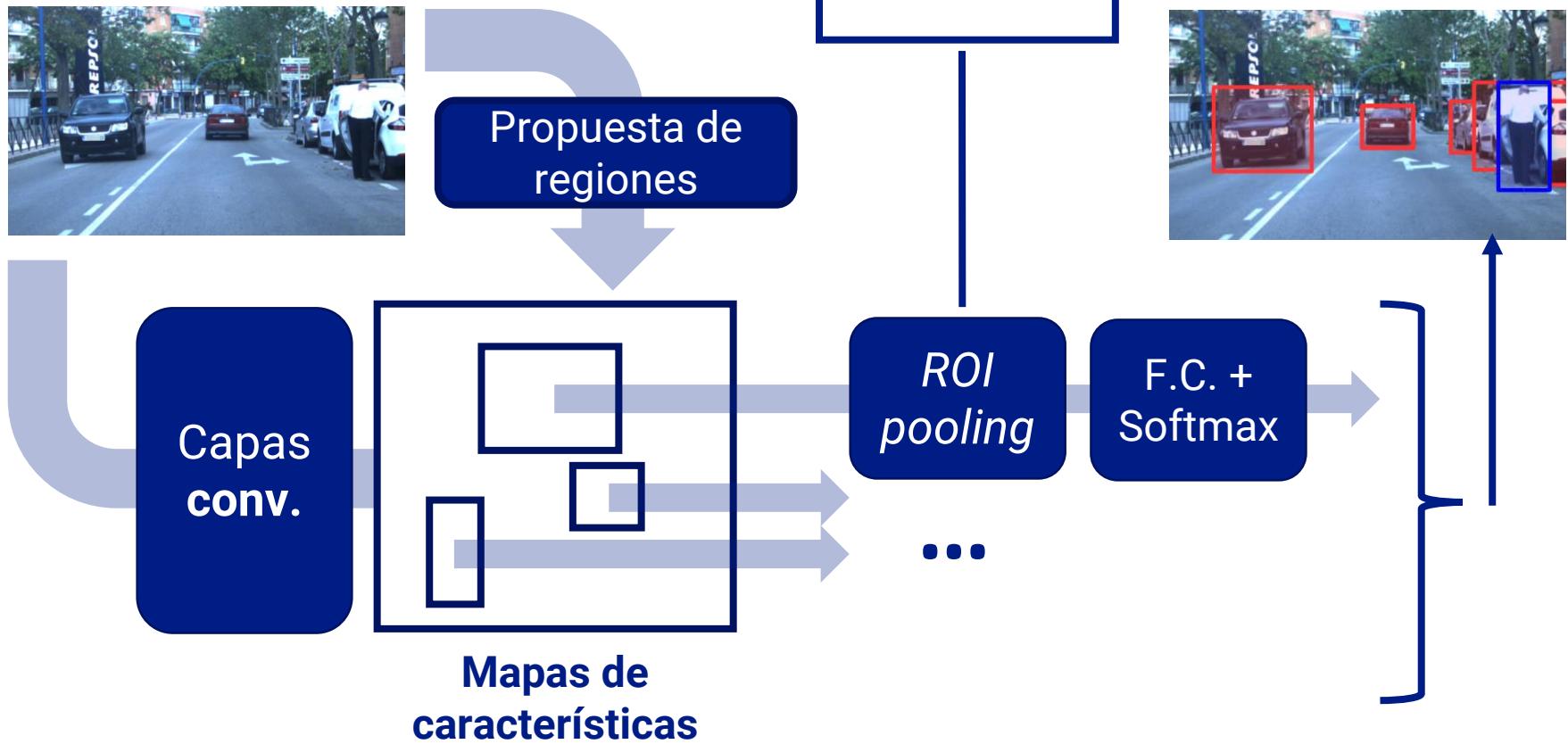
Deep learning

?

Fuente: [www.pyimagesearch.com](http://www.pyimagesearch.com), Adrian Rosebrock

# Detección: paradigma (Fast) R-CNN

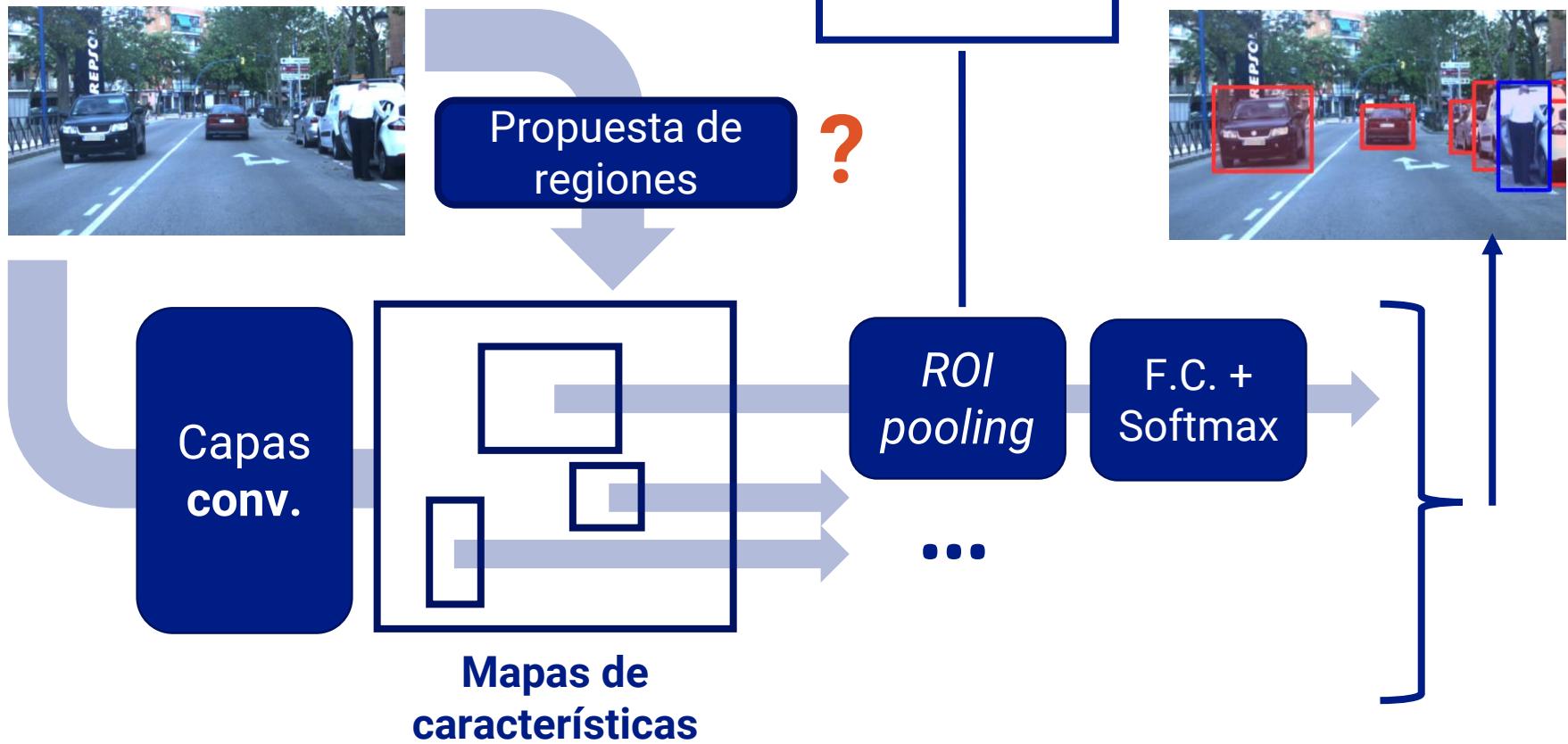
- CNN basadas en regiones
- Pérdida multi-tarea



R. Girshick, "Fast R-CNN," in Proc. IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.

# Detección: paradigma (Fast) R-CNN

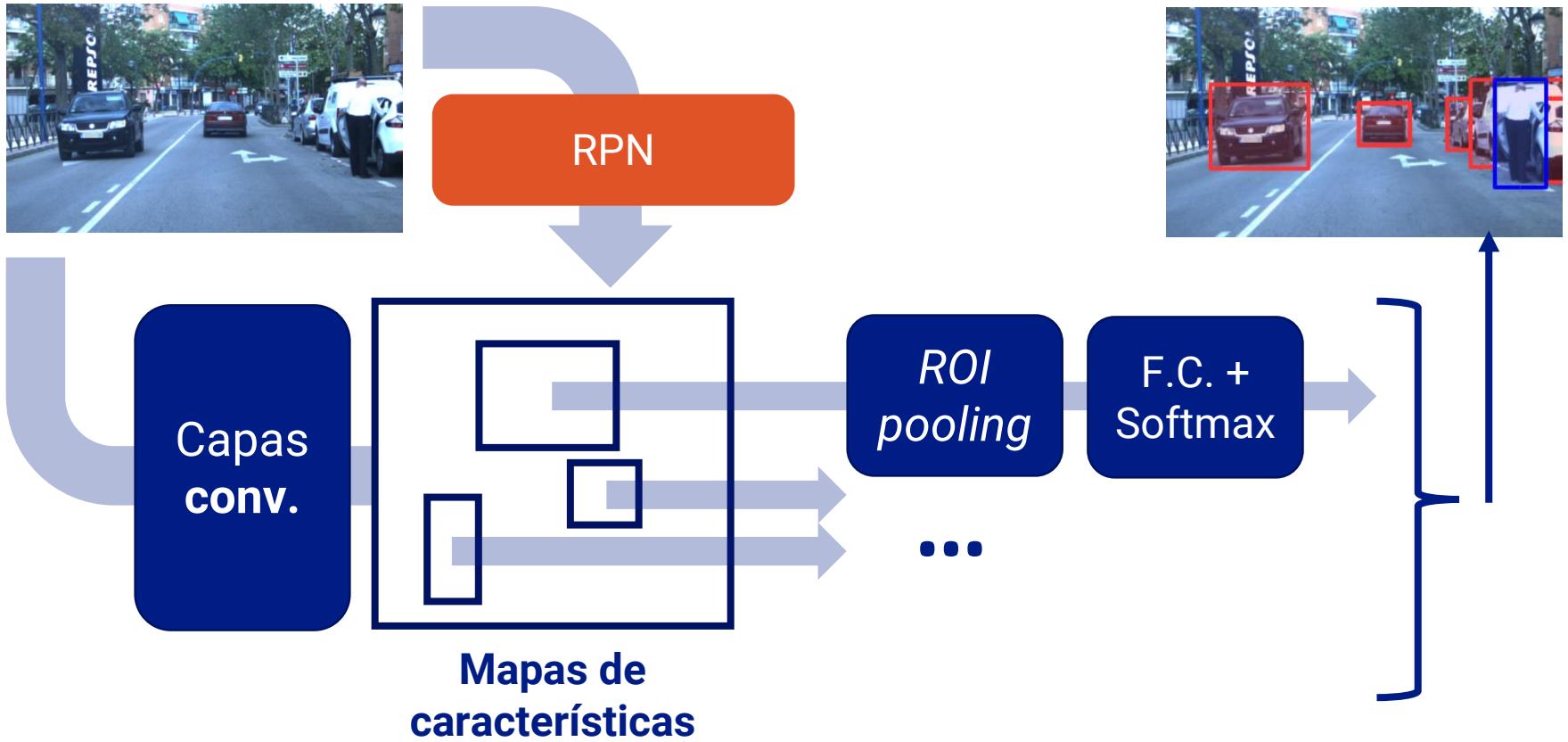
- CNN basadas en regiones
- Pérdida multi-tarea



R. Girshick, "Fast R-CNN," in Proc. IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.

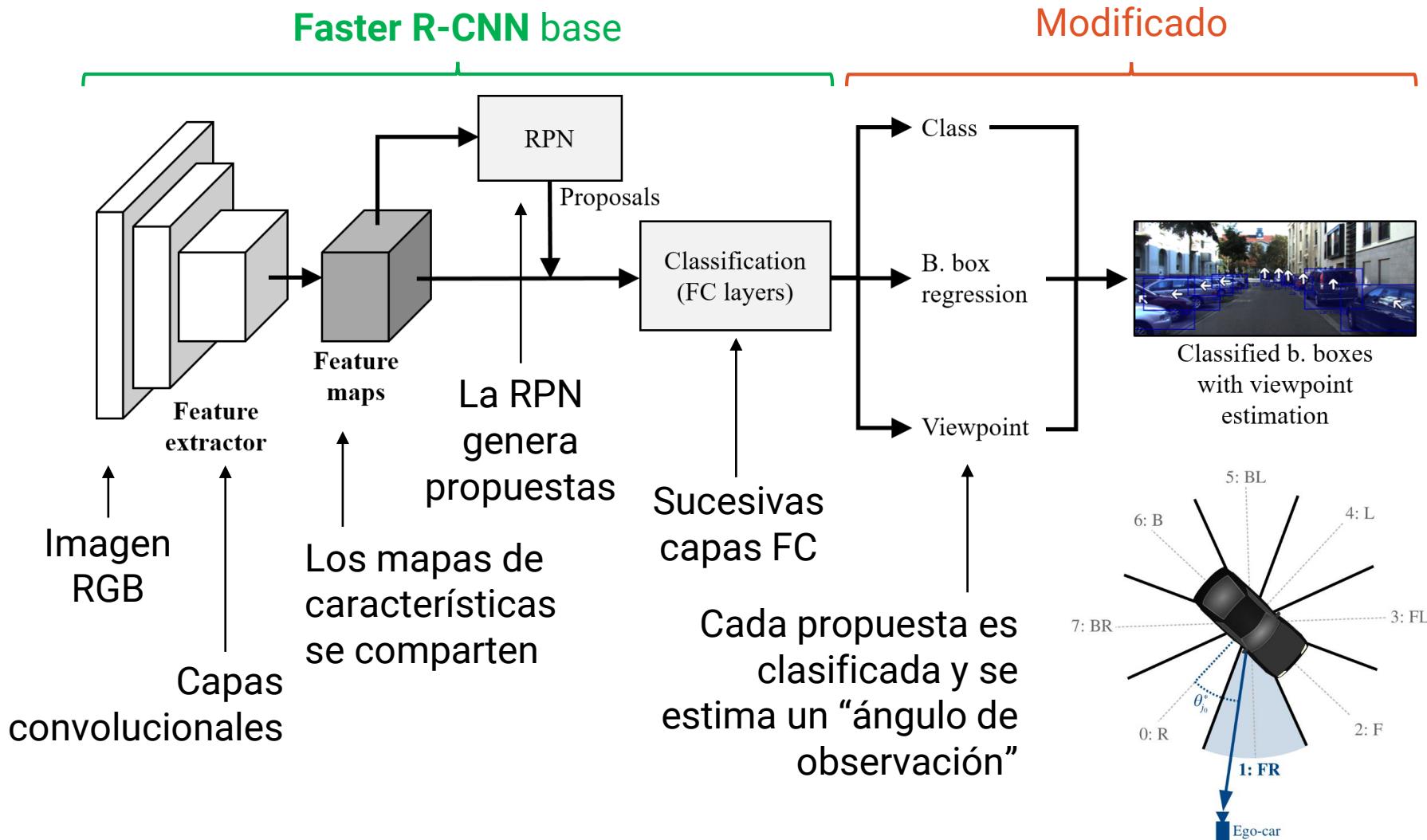
# Detección: paradigma Faster R-CNN

- “Red de propuesta de regiones” (**RPN**)
- Pequeña CNN que clasifica “cajas ancla” en objeto/no objeto



S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” in Advances in Neural Information Processing Systems (NIPS), 2015, pp. 91–99

# Detección y punto de vista (orientación)



C. Guindel, D. Martín, and J. M. Armingol, "Fast Joint Object Detection and Viewpoint Estimation for Traffic Scene Understanding," IEEE Intell. Transp. Syst. Mag., vol. 10, no. 4, pp. 74–86, 2018.

# Detección y punto de vista (orientación)

---



# Detección y punto de vista (orientación)

¿Cómo mejorar aún más el rendimiento?

## 1. Introducir información estéreo

Mapa de disparidades como un “cuarto canal”



Par  
estéreo

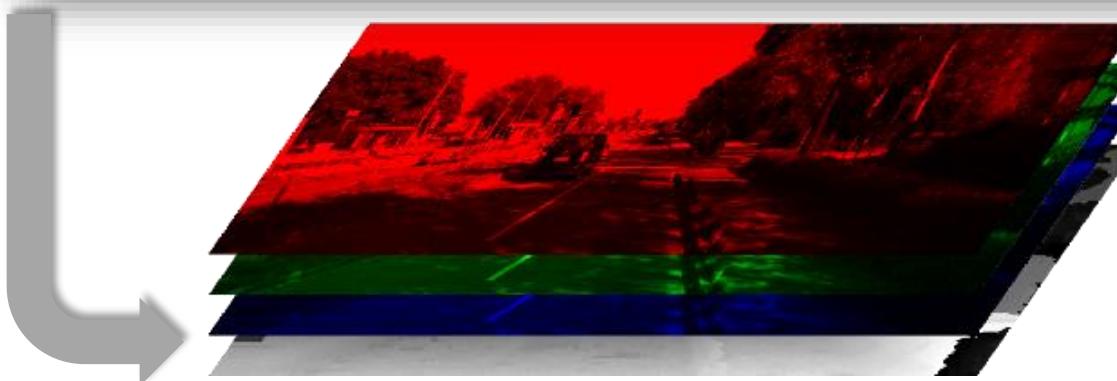
C. Guindel, D. Martín, and J. M. Armingol, "Stereo Vision-Based Convolutional Networks for Object Detection in Driving Environments," in Computer Aided Systems Theory - EUROCAST 2017, pp. 427–434.

# Detección y punto de vista (orientación)

¿Cómo mejorar aún más el rendimiento?

## 1. Introducir información estéreo

Mapa de disparidades como un “cuarto canal”



Aprox.  
+1 mAP

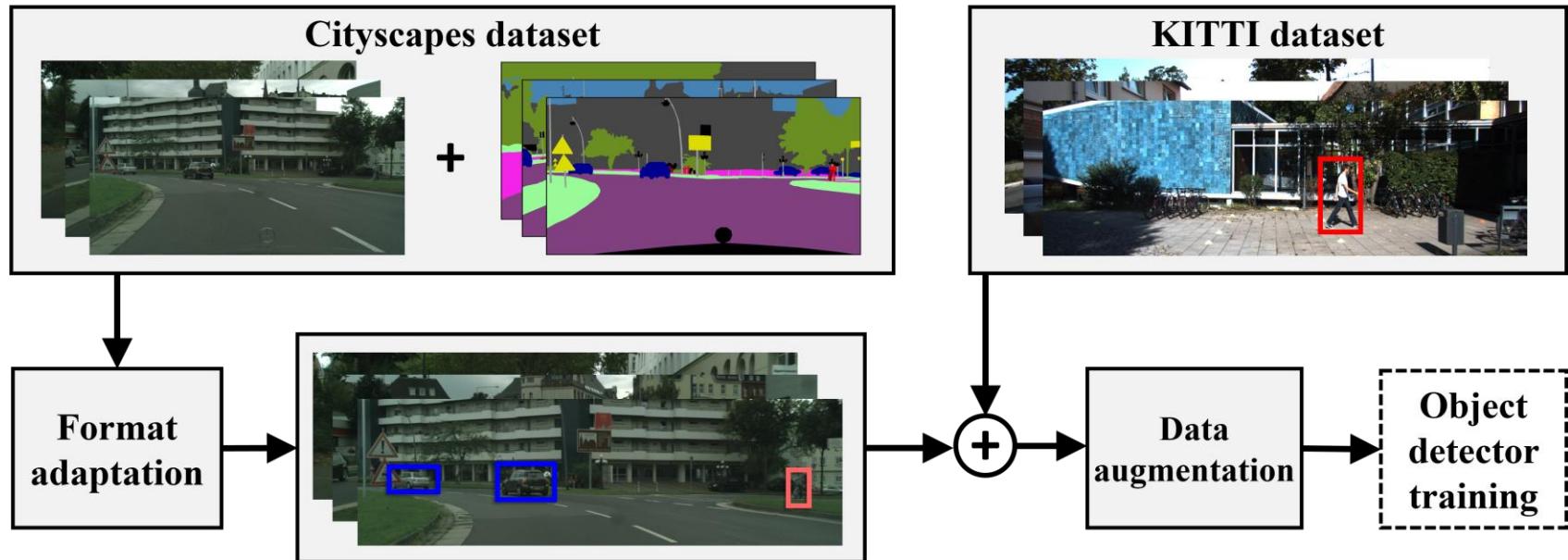
C. Guindel, D. Martín, and J. M. Armingol, "Stereo Vision-Based Convolutional Networks for Object Detection in Driving Environments," in Computer Aided Systems Theory - EUROCAST 2017, pp. 427–434.

# Detección y punto de vista (orientación)

¿Cómo mejorar aún más el rendimiento?

## 2. Combinar diferentes bases de datos

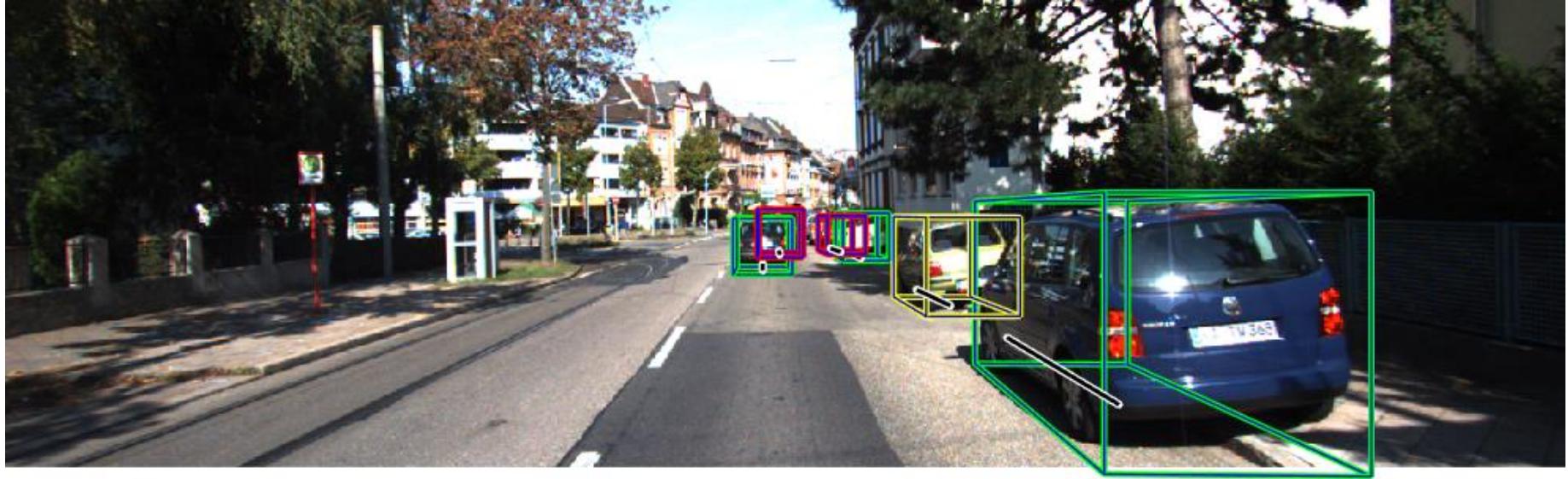
Mejora sustancial al introducir muestras adicionales



- BB DD de 3.712 imágenes: 65,98 % mAP
- Combinación 6.687 imágenes: **73,69% mAP**

C. Guindel, D. Martín, J. M. Armengol, and C. Stiller, "Analysis of the Influence of Training Data on Road User Detection," in Proc. IEEE International Conference on Vehicular Electronics and Safety (ICVES), 2018, pp. 21–26.

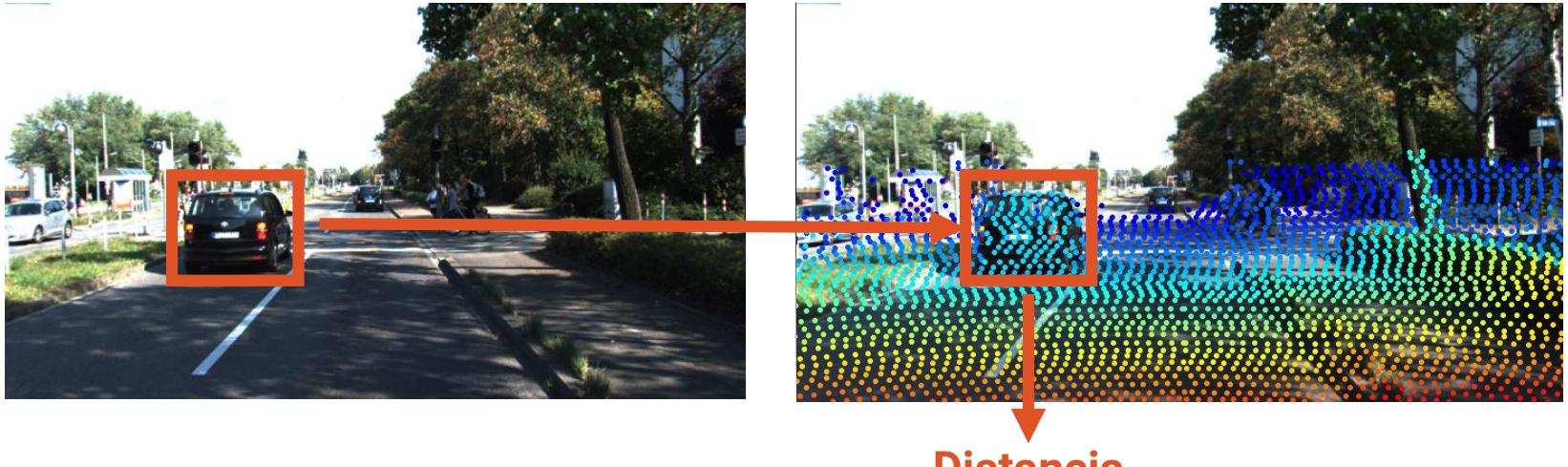
# Localización 3D



Fuente: KITTI dataset

- Coordenadas en la imagen:  $(u,v)$  ✓
  - ¿Coordenadas en el espacio  $(X,Y,Z)$ ?
  - ¿Tamaño de los objetos  $(l, w, h)$ ?
  - ¿Yaw  $(\theta)$ ?
- Se necesita información geométrica:
  - Estéreo: una sola modalidad, poco preciso
  - Lidar: alta precisión, menos información

# Localización 3D



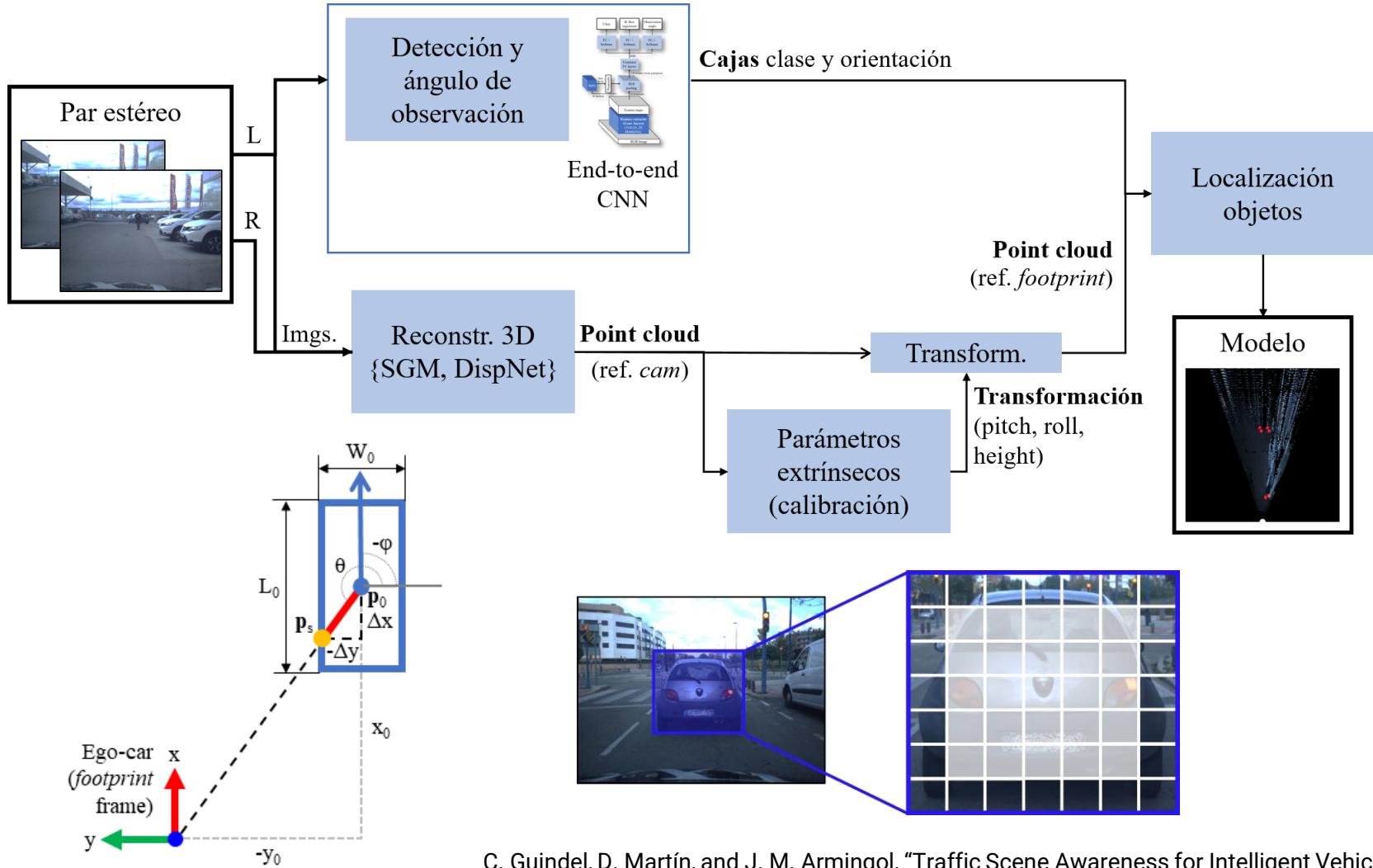
¿Punto más cercano?

¿Centroide?

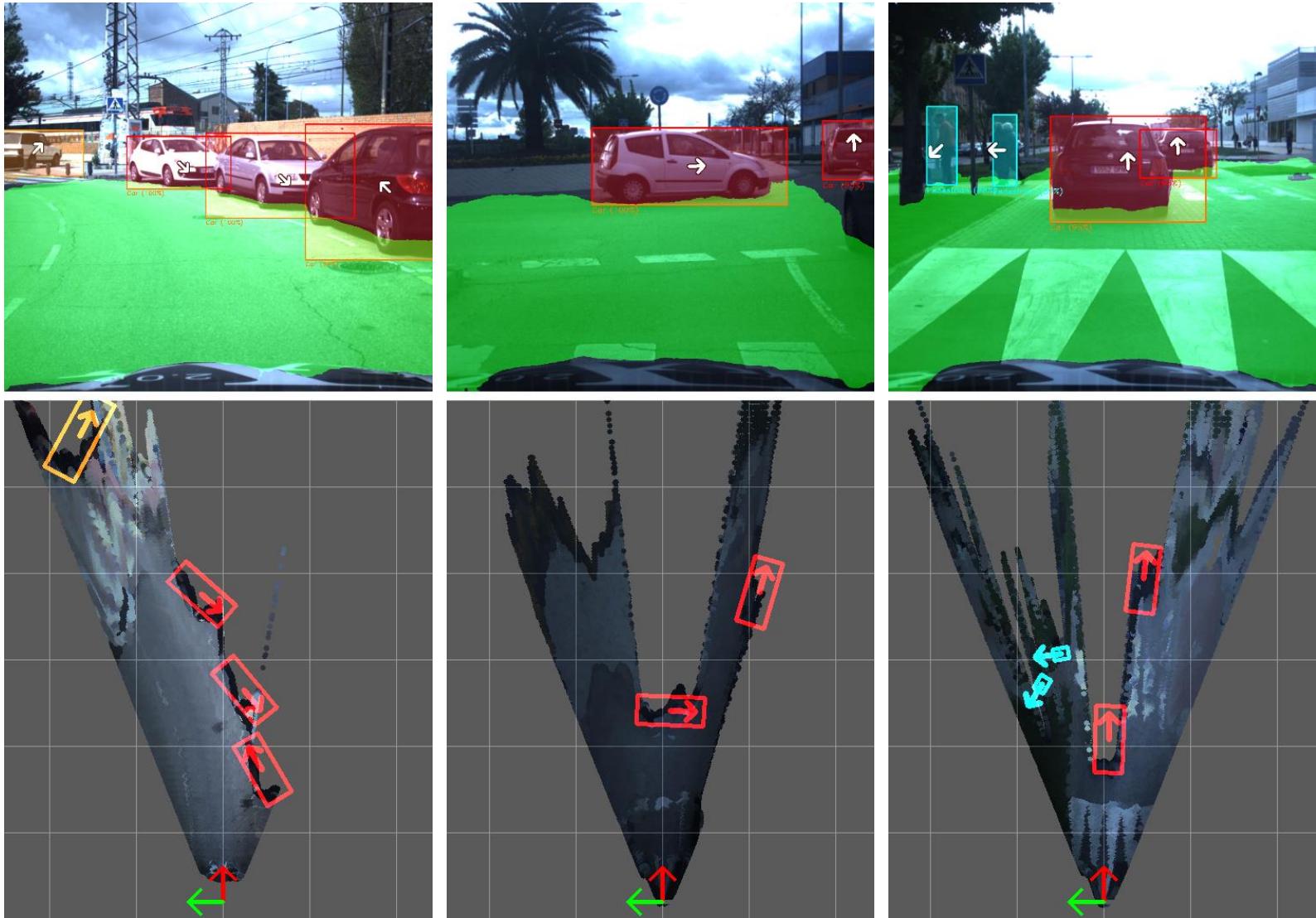
- ¿Y las dimensiones del objeto?
- ¿Y su orientación en coordenadas del mundo?

No hay medidas de distancia para las zonas no visibles por el sensor

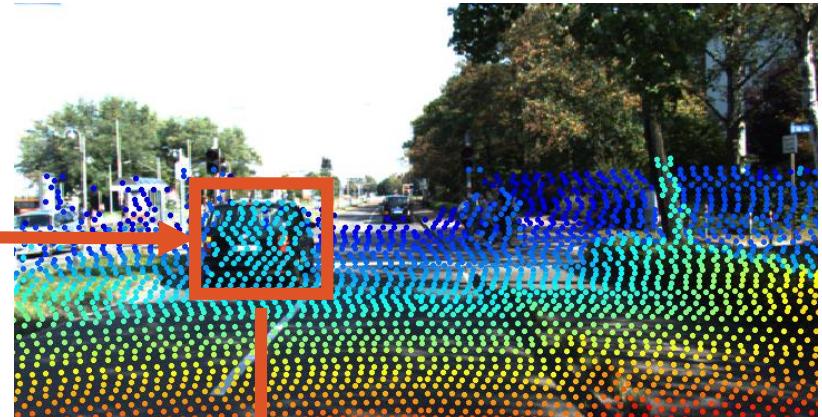
# Localización 3D: estéreo



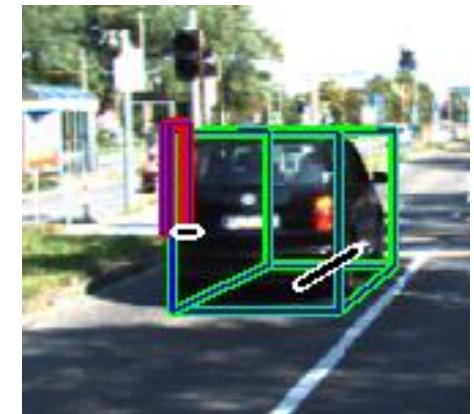
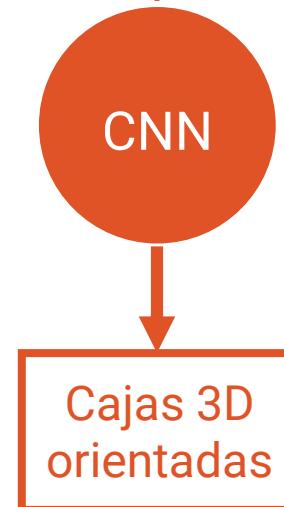
# Localización 3D: estéreo



# Localización 3D: alternativas

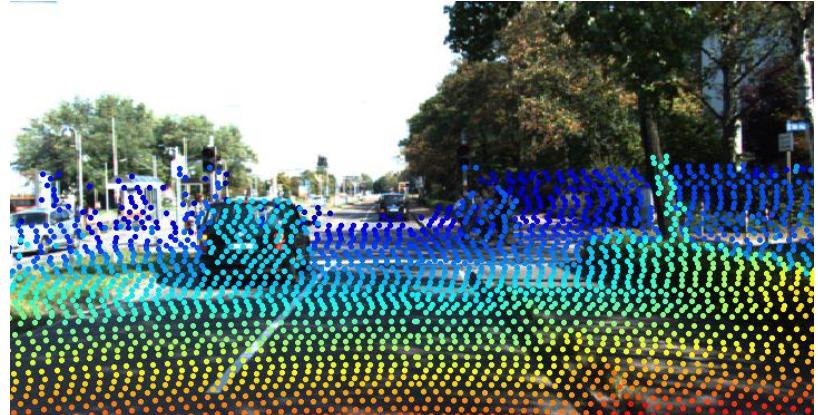


- Algunos métodos utilizan una red neuronal para estimar:
  - Puntos que pertenecen verdaderamente al objeto
  - Centro del objeto
  - Dimensiones del objeto
- Ej: Frustum PointNets

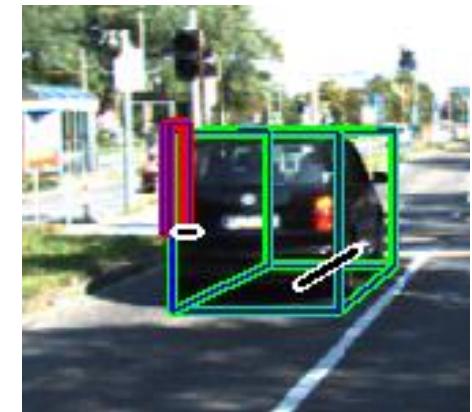
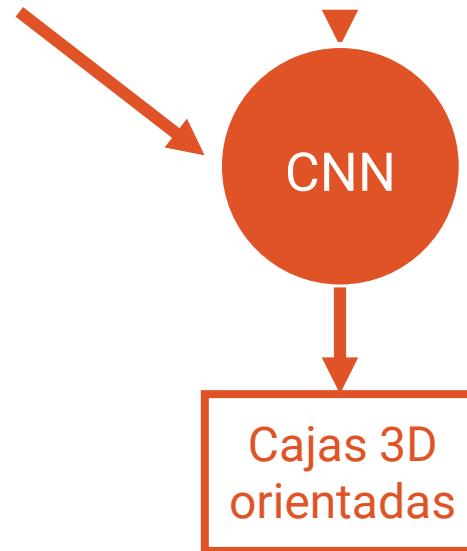


C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum PointNets for 3D Object Detection from RGB-D Data," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 918–927.

# Localización 3D: alternativas



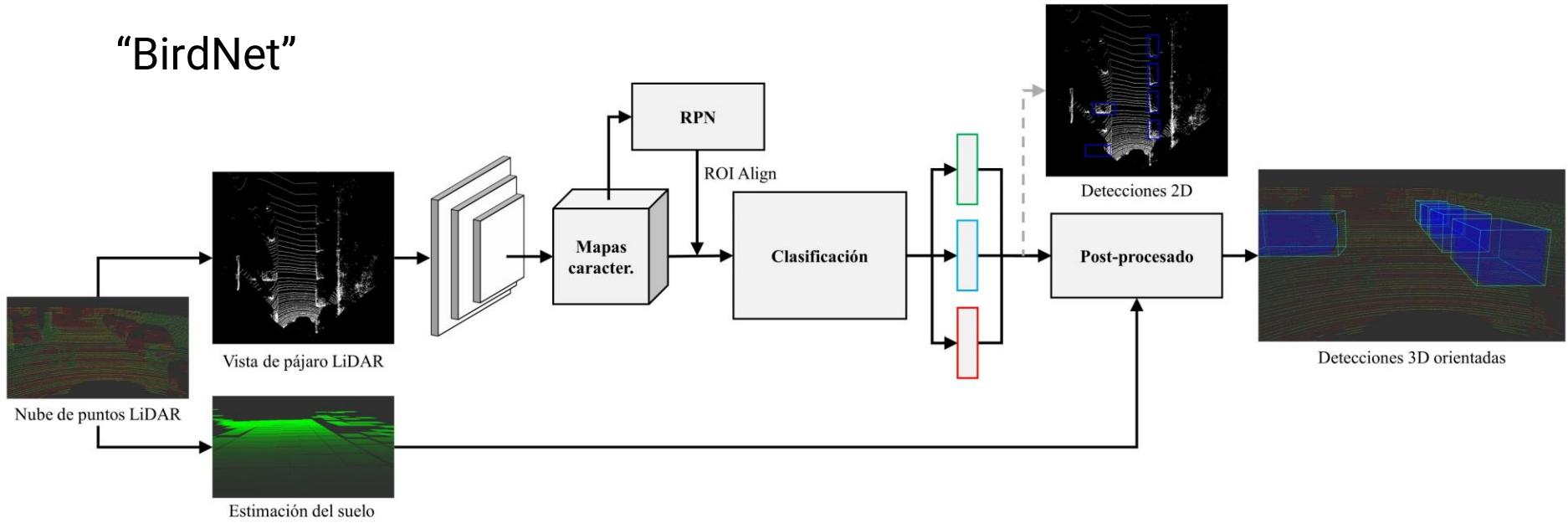
- Otros métodos estiman las detecciones directamente de la fusión de cámara y lidar
- El reto es la representación de la información
- Ej: MV3D



X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-View 3D Object Detection Network for Autonomous Driving," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6526–6534.

# Detección y localización 3D: lidar

## “BirdNet”

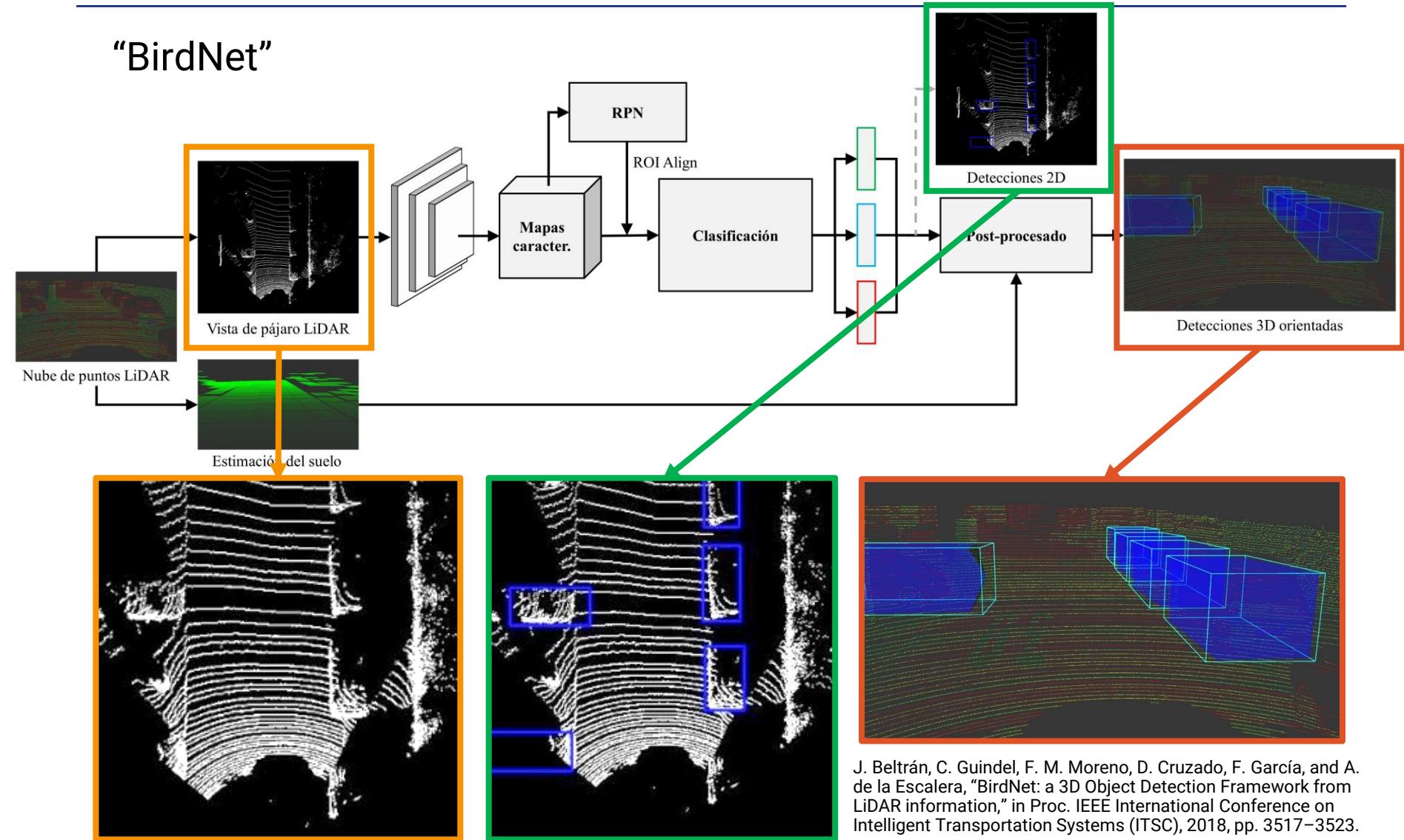


- Usa únicamente información del lidar (distancias)
- Esta información se representa de forma análoga a una imagen
- Detección y localización 3D simultáneas en 360°

J. Beltrán, C. Guindel, F. M. Moreno, D. Cruzado, F. García, and A. de la Escalera, “BirdNet: a 3D Object Detection Framework from LiDAR information,” in Proc. IEEE International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 3517–3523.

# Detección y localización 3D: lidar

## “BirdNet”



J. Beltrán, C. Guindel, F. M. Moreno, D. Cruzado, F. García, and A. de la Escalera, "BirdNet: a 3D Object Detection Framework from LiDAR information," in Proc. IEEE International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 3517–3523.

# Detección y localización 3D: lidar



## BirdNet: a 3D Object Detection Framework from LiDAR Information

Jorge Beltrán, Carlos Guindel, Francisco Miguel Moreno,  
Daniel Cruzado, Fernando García, and Arturo de la Escalera

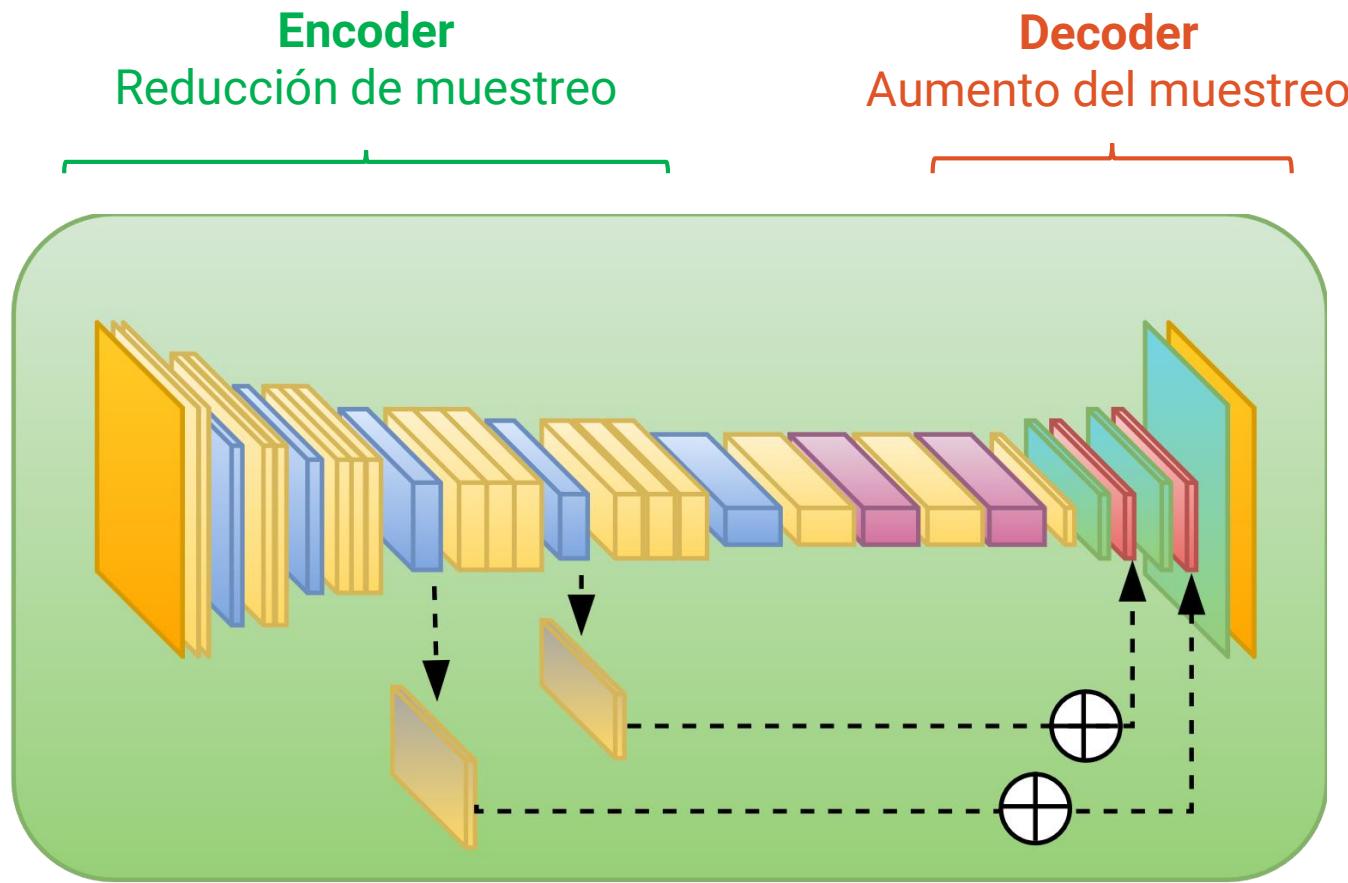
# Segmentación semántica



- Asignación de una clase a cada píxel
- Reconoce obstáculos pero no los diferencia entre sí
  - Los métodos “de instancias” sí (aplicando detección)
- Especialmente útil para elementos de la infraestructura:
  - Carretera, carriles, etc.

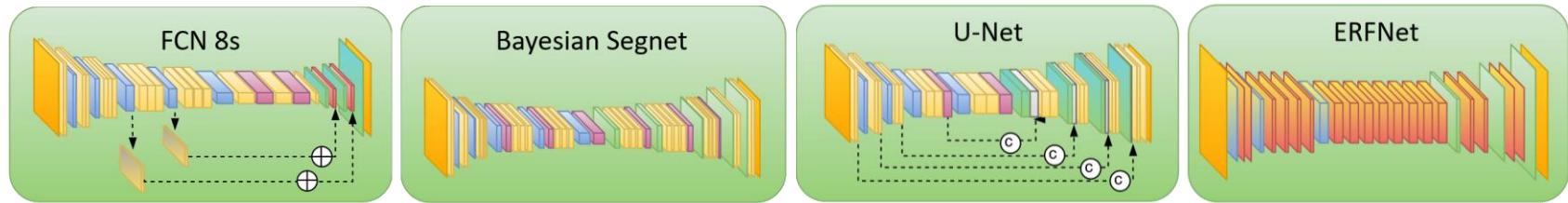
# Segmentación semántica

- Redes totalmente convolucionales



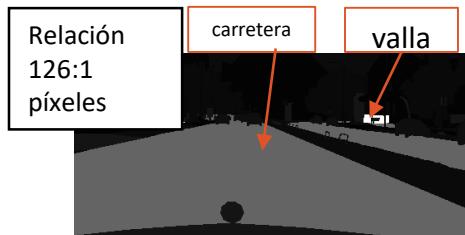
# Segmentación semántica

## Arquitecturas

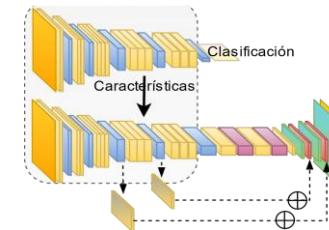


## Técnicas de entrenamiento

### Equilibrado de clases



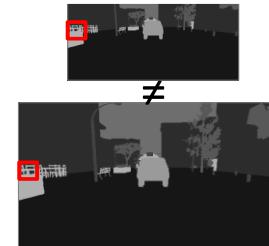
### Transferencia de aprendizaje



### Aumentado de datos



### Rescalado de la imagen



A. Barrera, C. Guindel, F. García, and D. Martín, "Análisis, evaluación e implementación de algoritmos de segmentación semántica para su aplicación en vehículos inteligentes," en XXXIX Jornadas de Automática, 2018, pp. 983–990.

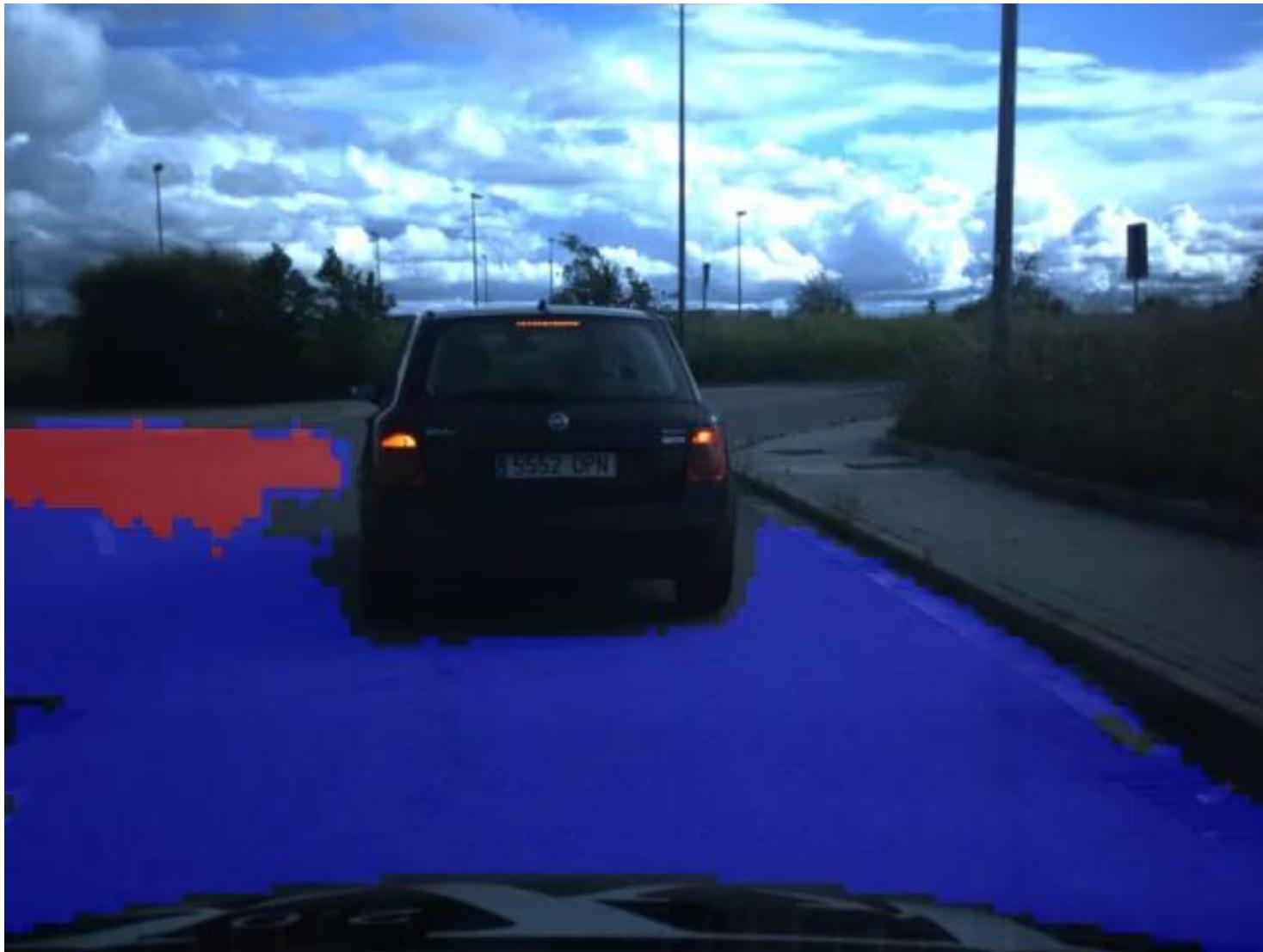
# Segmentación semántica

---



# Segmentación semántica: carriles

---

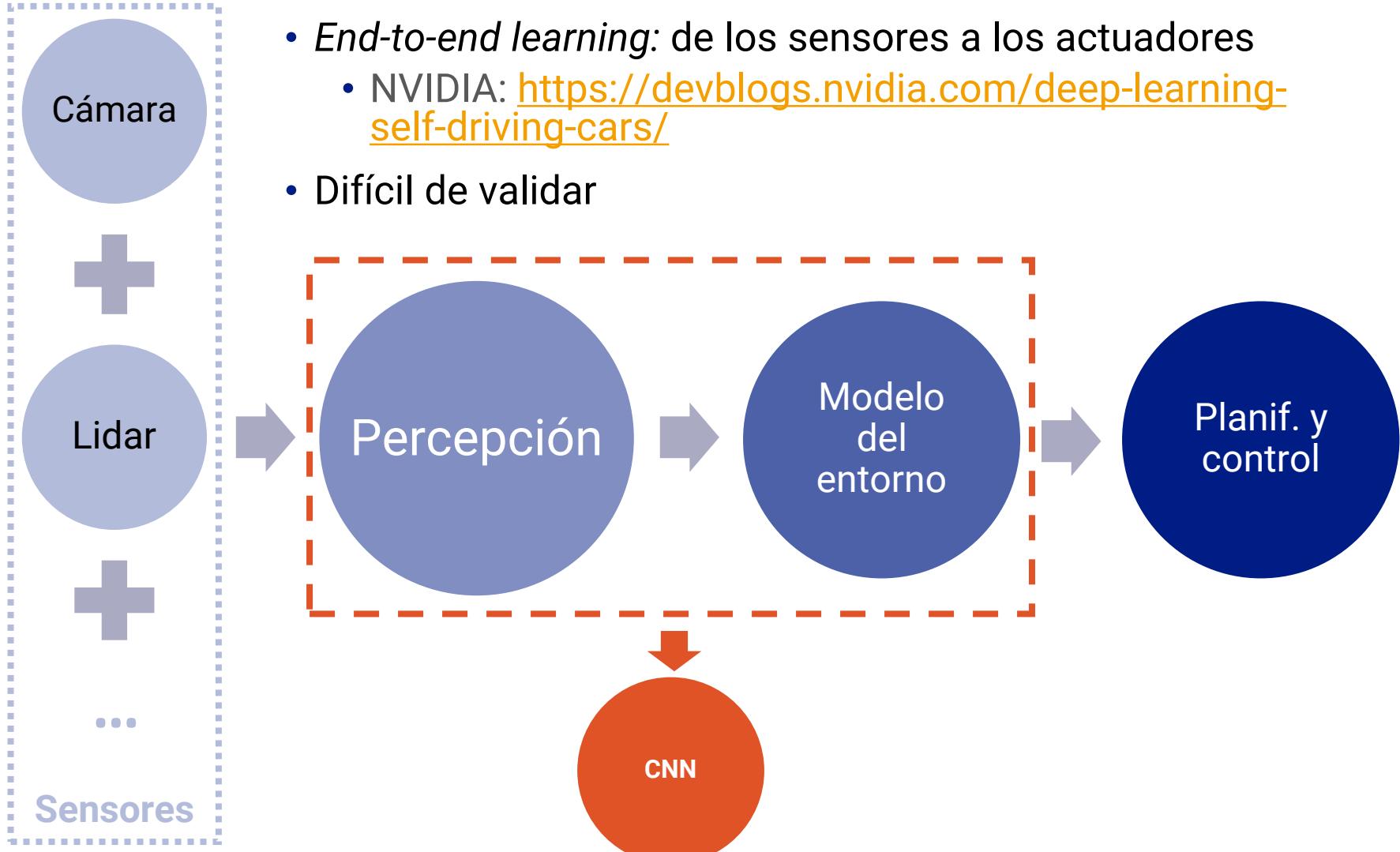


# Sistemas funcionales

---

<https://waymo.com/tech/>

# ¿Futuro?





¡MUCHAS GRACIAS!

cguindel@ing.uc3m.es



Webminar CEA  
24 de enero de 2019

