EVENTS: ReliablE in-Vehicle pErception and decisioN-making in complex environmenTal conditions

Javier Araluce, Tecnalia

12 de abril (Alcalá de Henares)

¿Quién es Javier Araluce?

Académico

- Grado en Ingeniería Electrónica y Automática Industrial
    - Automatización de un vehículo eléctrico
- Máster en Ingeniería Industrial: especialidad de Robótica y Percepción
    - Smart HMI for an autonomous vehicle
- Doctor en Electrónica: Sistemas Electrónicos Avanzados. Sistemas Inteligentes
    - Driver attention based on Deep Learning for a Smart Vehicle to Driver (V2D) interaction

Empresa

- Mettler-Toledo Safeline (Manchester)
- Asociación de bienes de equipo (Madrid)
- Tecnalia (Bilbao)

Funded by the European Union

Mi trabajo actual

- Líneas de investigación
  - Predicción de comportamientos
  - Detección de agentes 3D
  - Monitorización de ocupantes
    - Distracciones del conductor
    - Ocupación y estado de pasajeros
  - Percepción colaborativa
  - *Domain adaptation*

Funded by the European Union

Mi trabajo actual

- Proyectos
  - Financiación Publica Competitiva
    - **EVENTS**
    - Aware2All
    - AugmentedCCAM
    - SHELFY
  - Proyectos bajo contrato
    - DIVEC (Contratos con 3 clientes)
    - R3CAV

Funded by the European Union

Mi trabajo actual

- Alumnos
  - Alberto Justo (Alumno de doctorado)
    - Percepción colaborativa
    - Deep Learning
    - *Domain Adaptation*
    - Detección y Tracking

Funded by the European Union

EVENTS

- Objetivo: sistema de percepción y de decisión robusto que sea capaz de gestionar “eventos”
  - Eventos:
    - Interacción con otros agentes
    - Carreteras extrañas (obras…)
    - Baja visibilidad y cambios de clima
- Consorcio:
  - 11 empresas y centros de investigación
  - 8 países
- Datos:
  - 6,8 M € de presupuesto
  - 36 meses

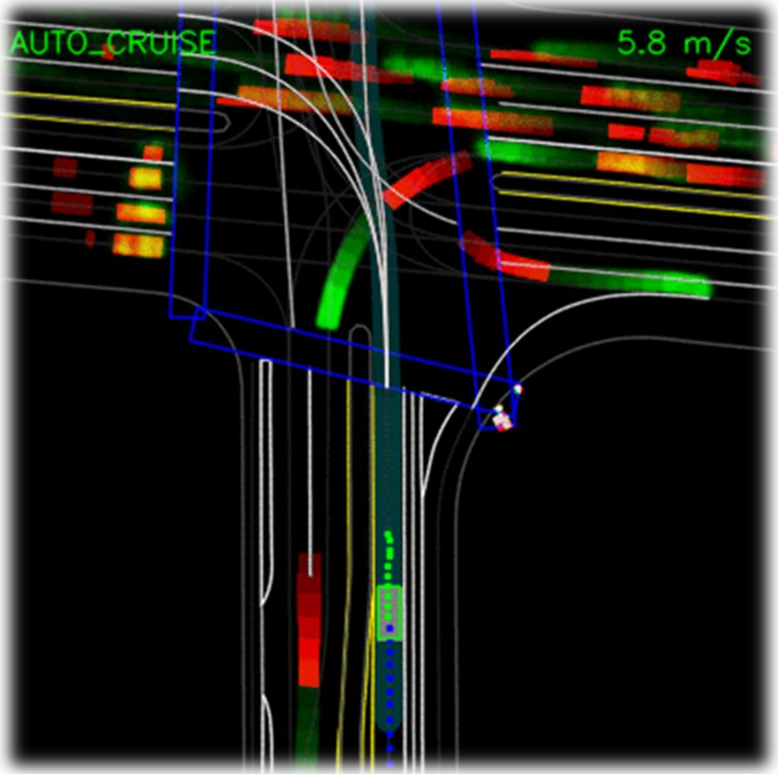
Papel de Tecnalia

- Desarrollos
  - Predicción de la trayectoria de otros agentes (WP3)
  - Percepción aumentada por V2X (WP3)
  - Sistema de decisión (WP4)
  - Sistemas ante fallos en el control (WP4)
- Experimentos
  - Maniobra de platooning en rotonda
  - Maniobra de evitación de una zona de obras
  - Estimación del comportamiento de los agentes en una intersección

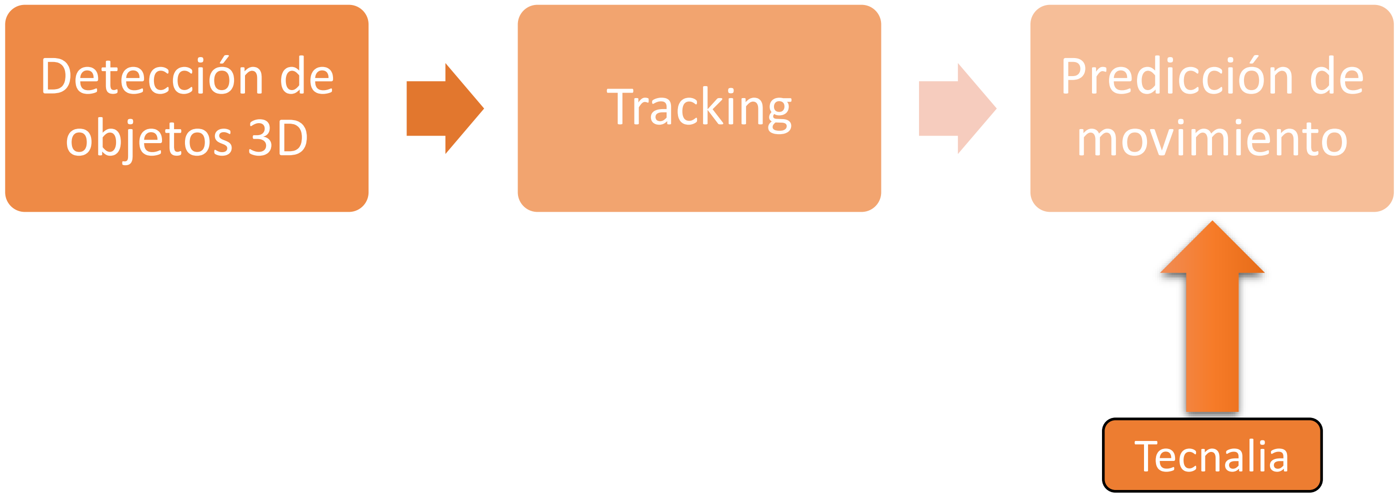
Funded by the European Union

Percepción

- Predicción de movimiento de otros agentes (Motion Prediction)
  - Predecir las trayectorias de los agentes cercanos
  - Permite obtener un comportamiento predictivo



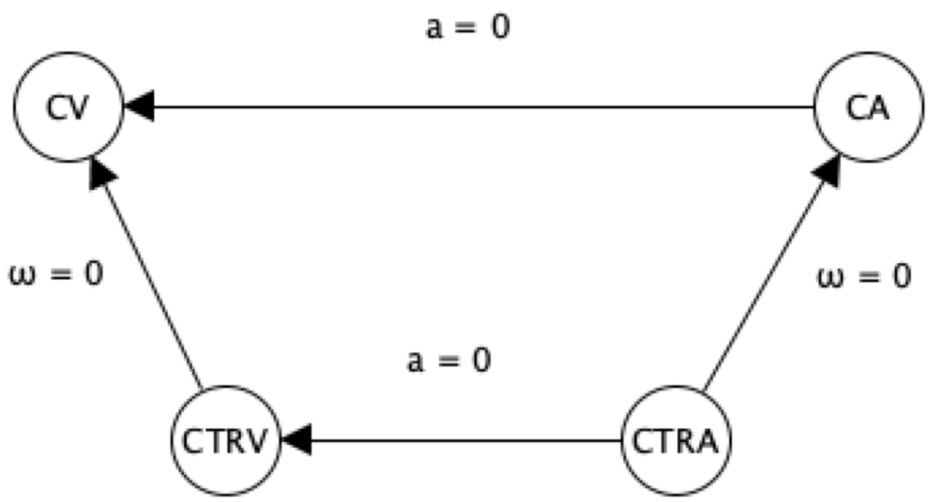
Sistema de percepción



Funded by the European Union

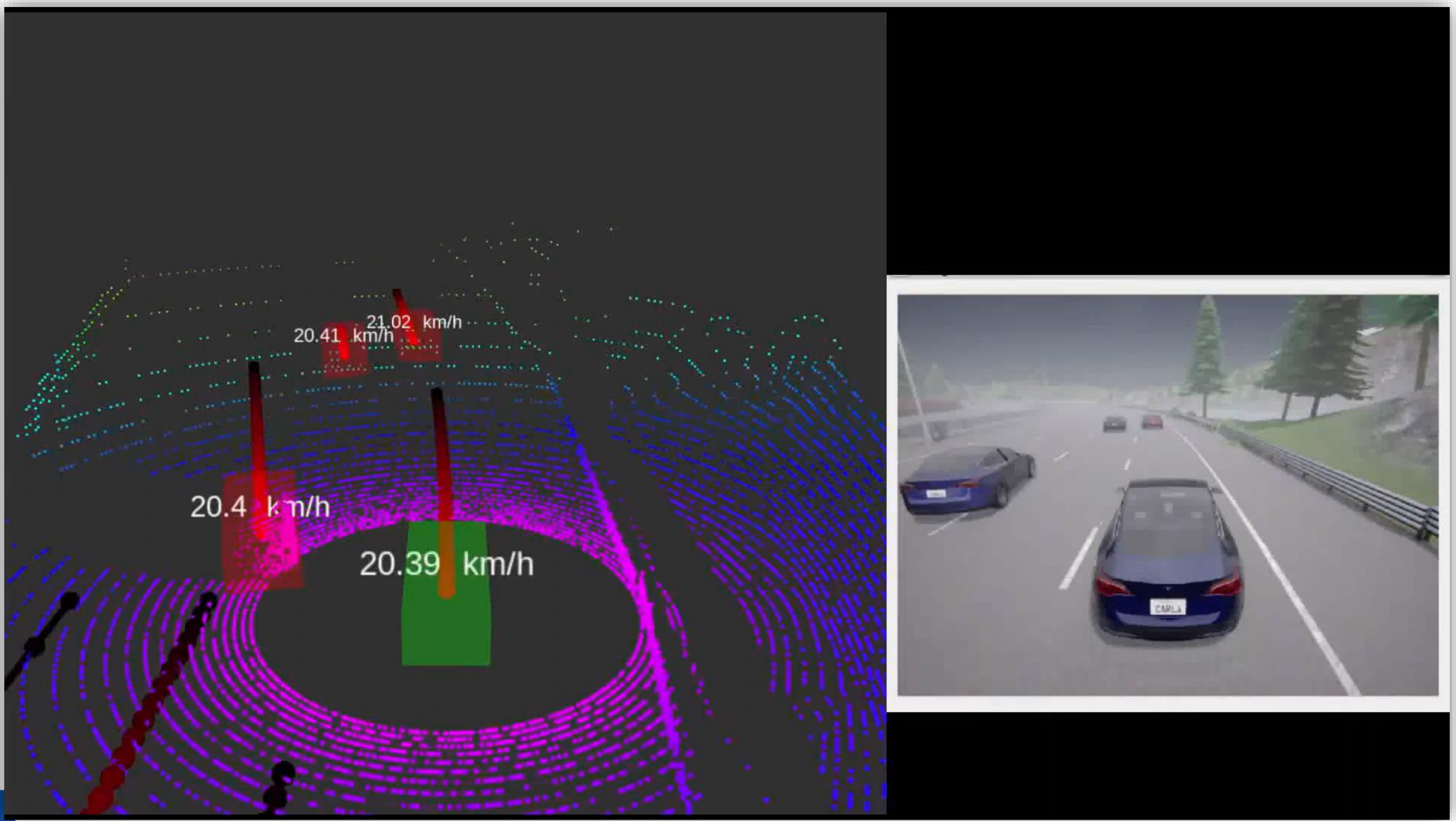
Motion Prediction

- Modelo físico clásico
- Utilizando el GT de nuestro simulador (CARLA)

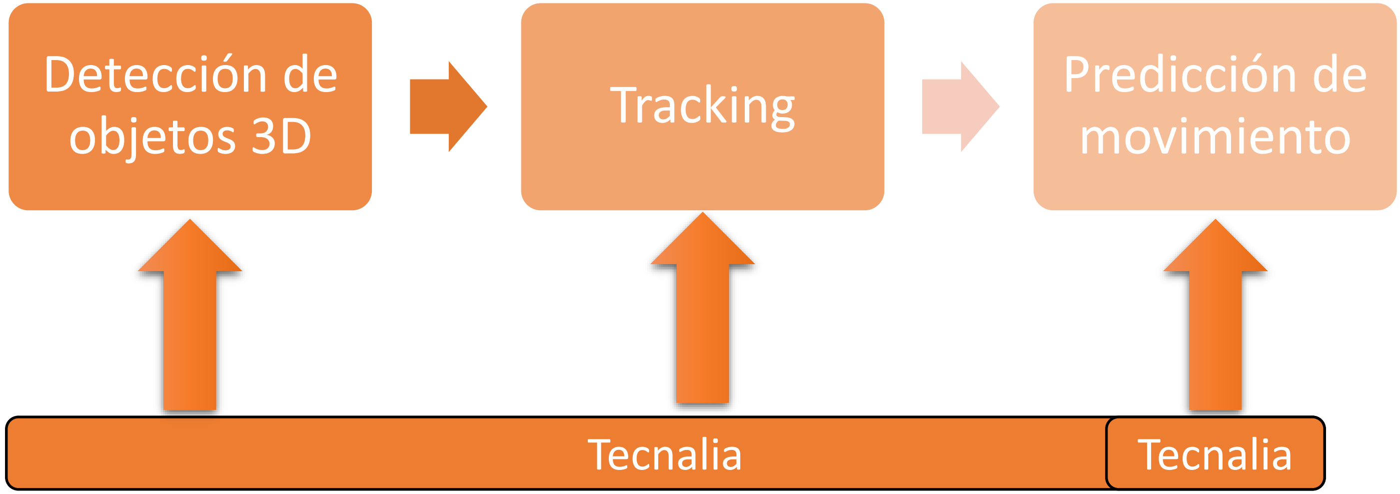


Funded by the European Union

Motion Prediction



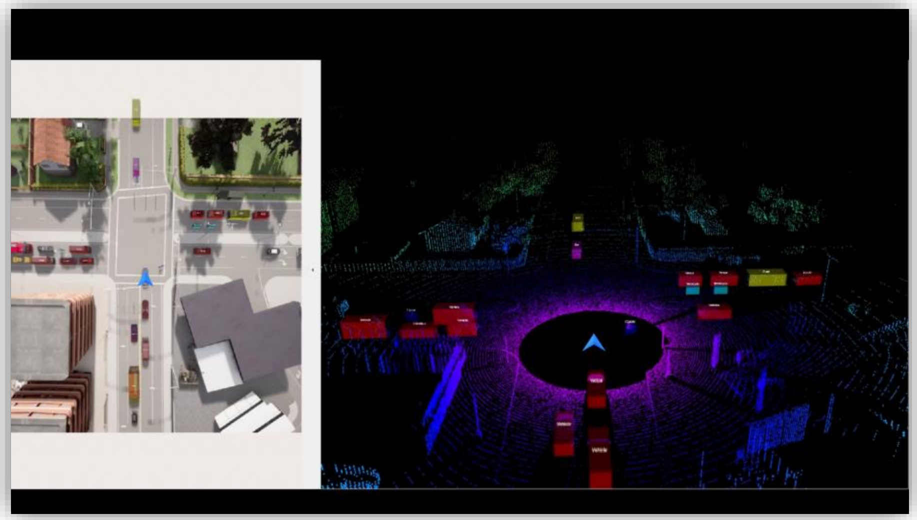
Sistema de percepción



Funded by the European Union

Detector 3D

- OpenPcdet
- PointPillars [1]
- Entrenado en dato sintético (SHIFT [2])



[1]: Lang, Alex H., et al. "Pointpillars: Fast encoders for object detection from point clouds." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
[2]: Sun, Tao, et al. "SHIFT: a synthetic driving dataset for continuous multi-task domain adaptation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Funded by the European Union

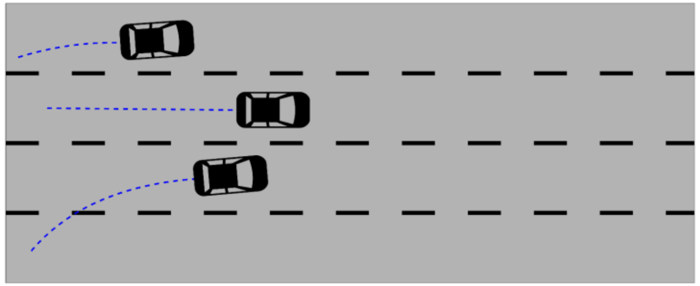
Multi Object Tracking

- Algoritmo SORT
  - Algoritmo húngaro (asociaciones)
  - Filtro de Kalman (predicciones)

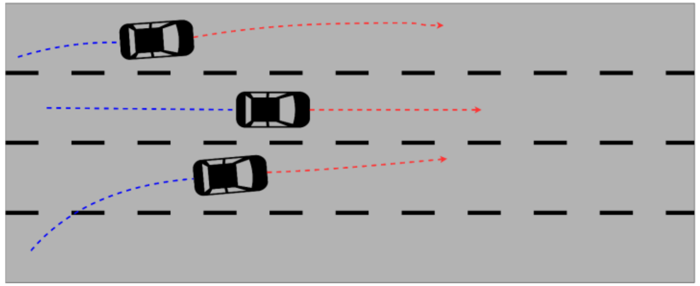
Funded by the
European Union

Motion Prediction

- Modelo físico clásico
- Utilizando el sistema de percepción



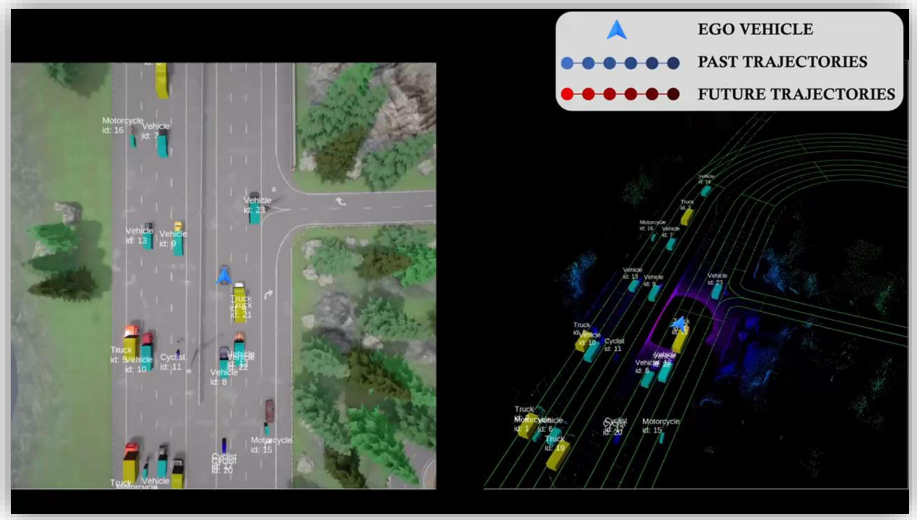
3D Object Detection (PointPillars) → Multi Object Tracking → Past Object trajectories (20 frames = 2 seconds) —[x, y, \theta, v a, \omega]→ CTRV/CTRA model → Future Object trajectories (30 frames = 3 seconds)



Funded by the European Union

Motion Prediction

- Modelo físico clásico
- Utilizando el sistema de percepción



Funded by the European Union

Motion Prediction

Crat-Pred [1]

- Modelo social
- No depende de mapa
- 6 trayectorias
- 200 ms de inferencia (RTX 4000)
- Licencia: Creative Commons Attribution-NonCommercial 4.0 International

HiVT [2]

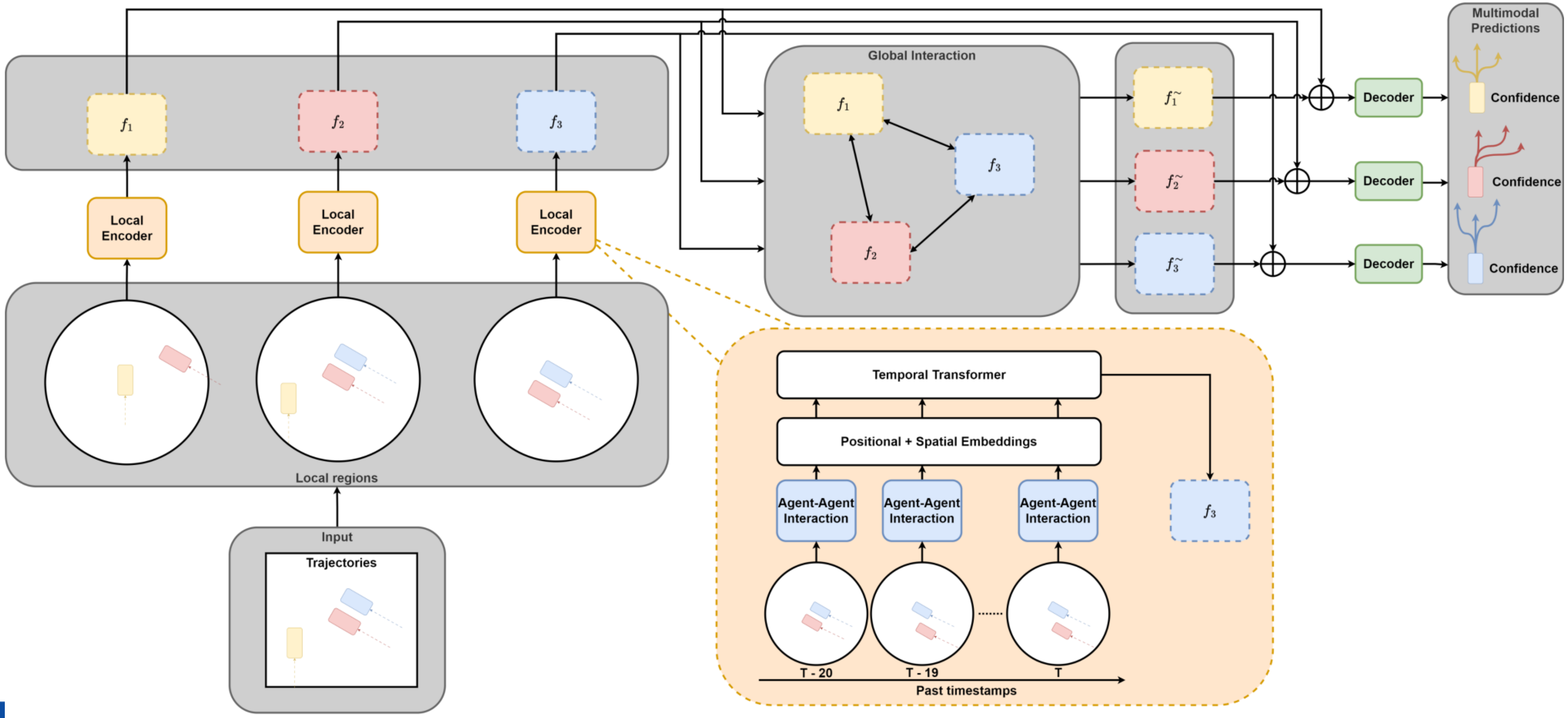
- Modelo social con mapa
- 6 trayectorias con confidencia
- 80 ms de inferencia (RTX 4000)
- Licencia: Apache

[1]: Schmidt, Julian, et al. "Crat-pred: Vehicle trajectory prediction with crystal graph convolutional neural networks and multi-head self-attention." *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022.
[2]: Zhou, Zikang, et al. "Hivt: Hierarchical vector transformer for multi-agent motion prediction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Funded by the European Union

Motion Prediction

- HiVT



Funded by the European Union

Motion Prediction

- Información social: relación entre agentes
- Resultados de validación en Argoverse 1
  - minADE: mínimo error medio
  - minFDE: mínimo error final
  - MR: % de rutas con un desplazamiento final mayor a 2 metros

Model	Map	minADE ↓	minFDE ↓	MR (%) ↓
HiVT-64 [7]	✓	0.69	1.04	0.10
HiVT-128 [7]	✓	**0.66**	**0.96**	**0.09**
Crat-Pred [28]	✗	0.85	1.44	0.17
HiVT-64 (ours)	✗	0.76	1.24	0.14

Funded by the European Union

Motion Prediction

Información de mapa

Pros

- Soluciones más precisas
- Contexto espacial
- Tienes una ligera información del futuro. En casos normales

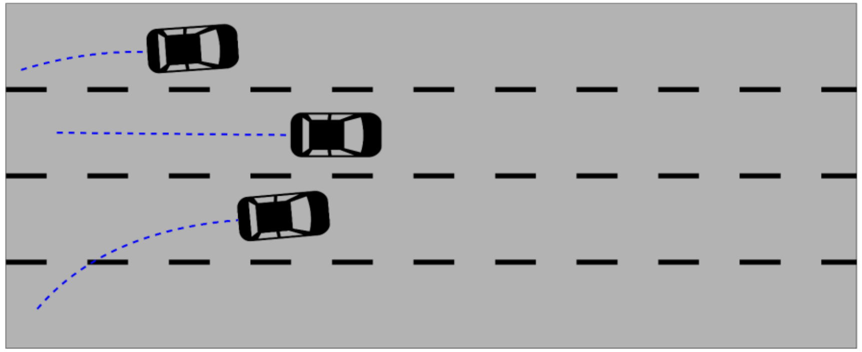
Contras

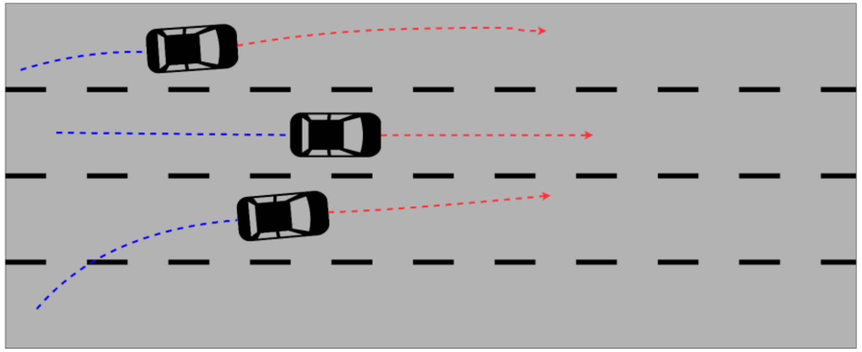
- Poco escalable:
  - Muchos formatos
- Información difícil de manejar

Funded by the European Union

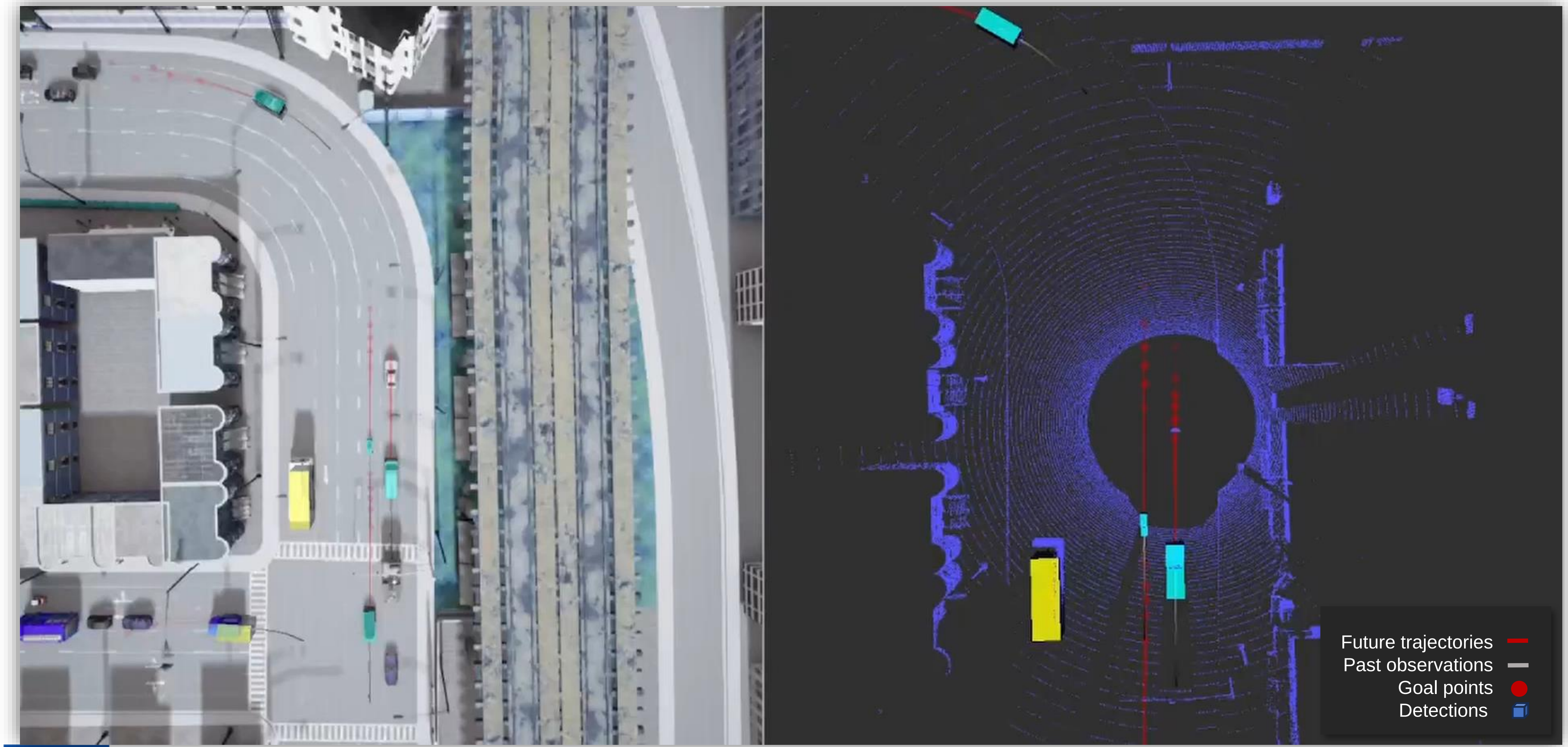
Motion Prediction

LiDAR-based 3D Object Detection (PointPillars) → Multi Object Tracking (tracking by detection) → Past Object trajectories (20 frames = 2 seconds) → HiVT map-less → Future Object trajectories (30 frames = 3 seconds)



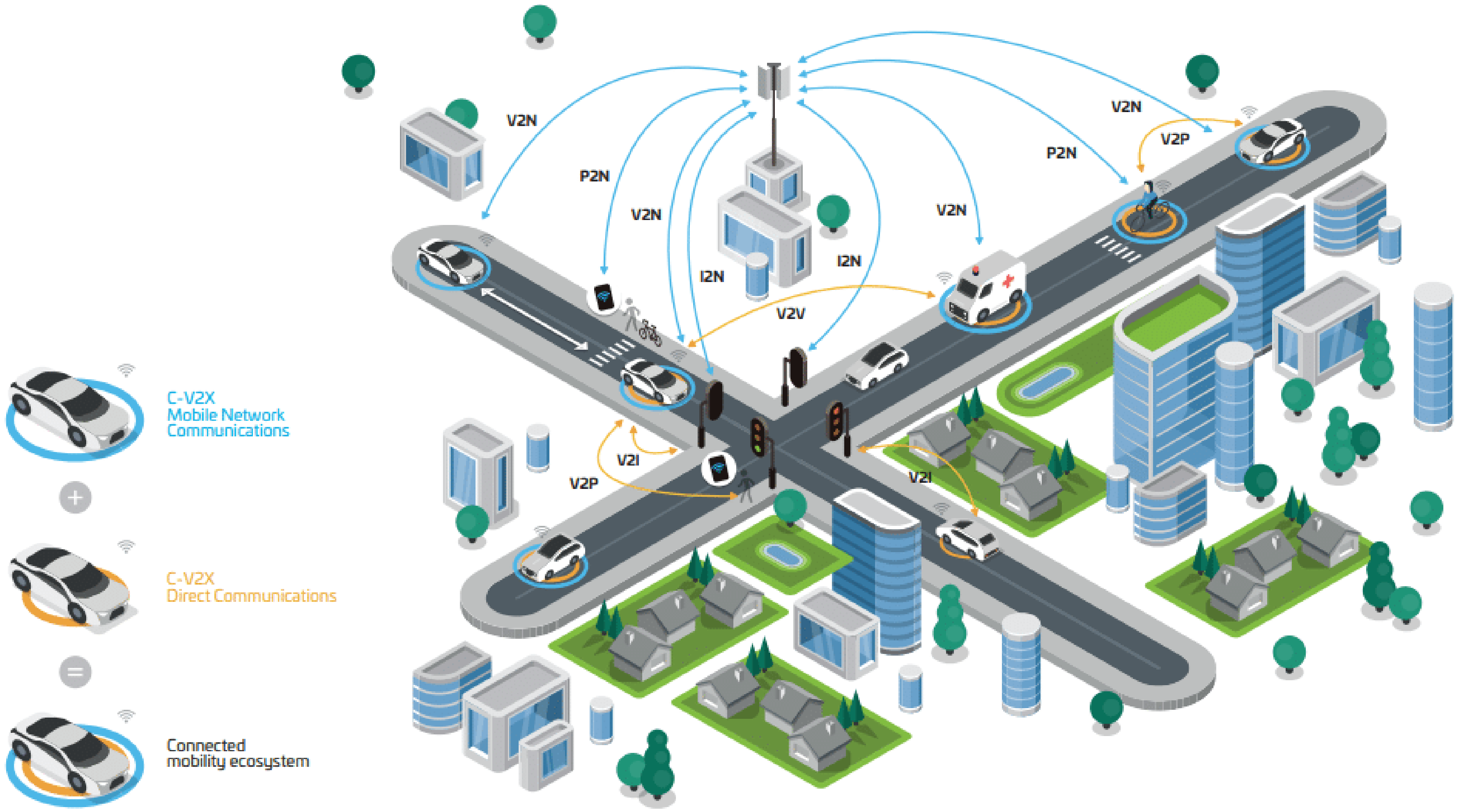


Motion Prediction



Funded by the European Union

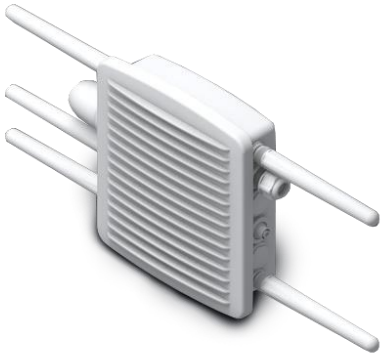
Percepción aumentada por V2X



Funded by the
European Union

Percepción aumentada por V2X

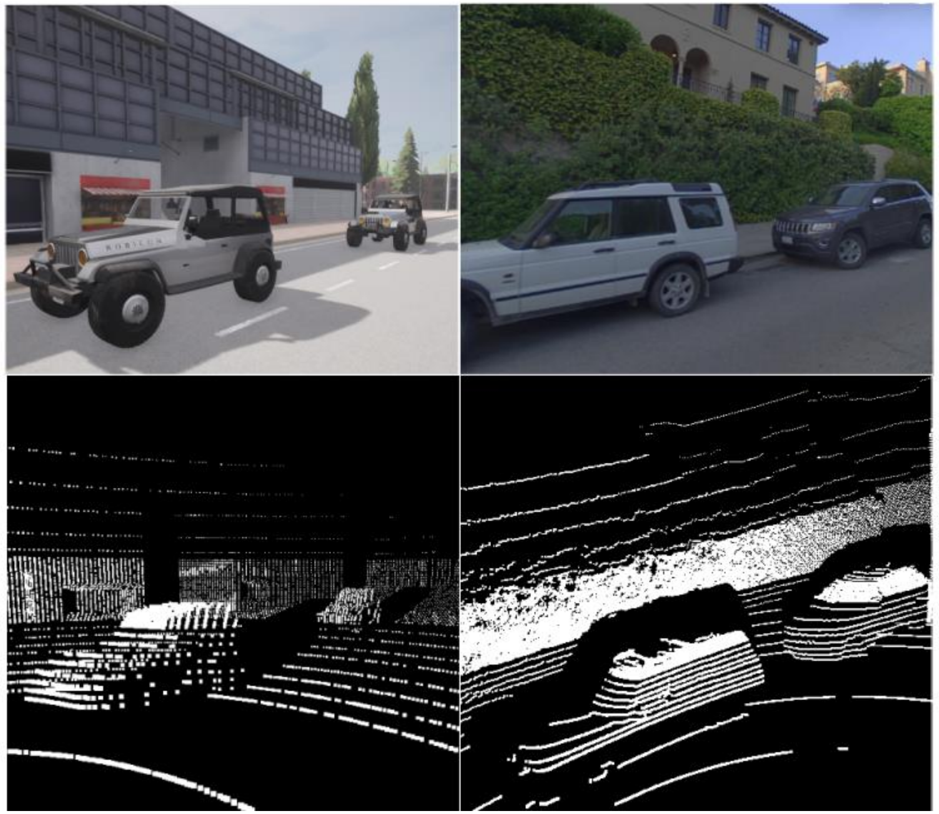
- Hardware:
  - OBU: OnBoard Unit
  - RSU: Road Side Unit
- Software:
  - ETSI messages
    - CAM: Cooperative Awareness Message
    - CPM: Collective Perception Message



Funded by the
European Union

Simulación realista

- El LiDAR en CARLA no es "realista"
- Los sistemas basados en LiDAR no son trasladables a otros entornos
- Problema de *Domain Adaptation*



- Alberto Justo , Javier Araluce , Mario Rodríguez-Arozamena , Leonardo Gonzalez and Sergio Díaz, "SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception", in 2024 IEEE Intelligent Vehicles Symposium (IV), IEEE, June 2024.

Funded by the European Union

SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception

- CARLA genera la nube por Ray tracing.
- Depende de:
  - Ángulo de azimuth (θ)
  - Ángulo polar ($\emptyset$)
  - Distancia obtenida por ray casting (r_i)
- En CARLA son ideales
- Distribuciones uniformes y consistentes
- Esto no es realista

$$p_i = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = r_i \begin{pmatrix} \cos(\phi)\cos(\theta) \\ \cos(\phi)\sin(\theta) \\ \sin(\phi) \end{pmatrix}$$

Funded by the European Union

SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception

- Intensidad se calcula con Beer-Lambert Law
  - I: intensidad
  - I_o: intensidad inicial antes de entrar al material
  - a: coeficiente del medio
  - d: espesor del medio que la luz atraviesa

$$\frac{I}{I_0} = e^{-a \cdot d}$$

- CARLA lo codifica utilizando el RGB en una imagen virtual

Funded by the European Union

SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception

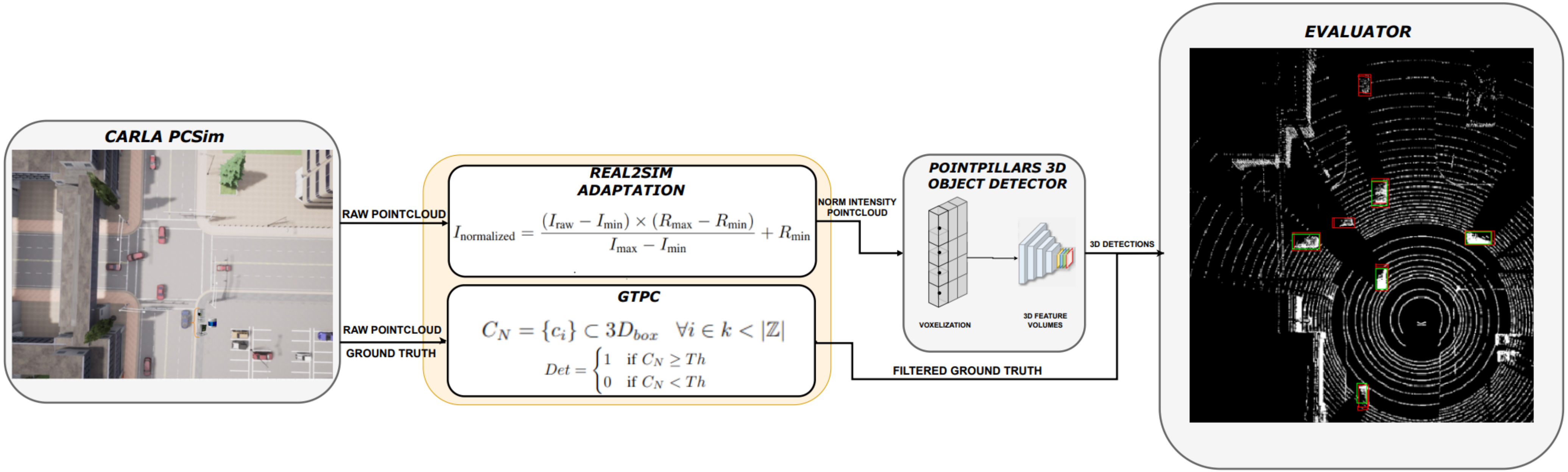
- PCSim [1]: simula diferentes LiDARs del mercado
  - Ángulo de azimuth (θ)
  - Ángulo polar (Ø)
  - Distancia obtenida por ray casting (r_i)
  - La intensidad es simulada en valores de *drop-off*
  - La intensidad en cada sensor tiene unas cotas diferentes
- Normalización de la Intensidad

$$I_{\text{normalized}} = \frac{(I_{\text{raw}} - I_{\text{min}}) \times (R_{\text{max}} - R_{\text{min}})}{I_{\text{max}} - I_{\text{min}}} + R_{\text{min}}$$

[1] X. Cai et al, "Analyzing infrastructure lidar placement with realistic lidar simulation library," in 2023 IEEE International Conference on Robotics and Automation (ICRA), IEEE, May 2023.

Funded by the European Union

SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception



Funded by the European Union

SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception

- Datasets

Dataset	LiDAR	Channels	vFOV(º)	Range(m)	Frames	BBoxes
KITTI [29]	Velodyne HDL-64	64	27	120	7.4k	80k
NuScenes [17]	Velodyne HDL-32	32	41	100	390k	1.4M
Pandaset [12]	Hesai Pandar64	64	40	200	16k	-

Dataset	IoU	mAP@0.7
KITTI	0.7	0.39
NuScenes	0.84	0.68
Pandaset	0.88	0.71

Funded by the European Union

SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception

- Resultados

Dataset	PCSim LiDAR	DA	IoU	mAP@0.7
KITTI	HDL-64	-	0.67	0.34
	HDL-64	Real2Sim	**0.71**	**0.43**
NuScenes	HDL-32	-	0.77	0.59
	HDL-32	Real2Sim	0.77	0.59
	CARLA-32	-	0.76	0.54
	CARLA-32	Real2Sim	0.76	0.54
Pandaset	Pandar64	-	0.86	0.65
	Pandar64	Real2Sim	**0.88**	**0.73**

Funded by the European Union

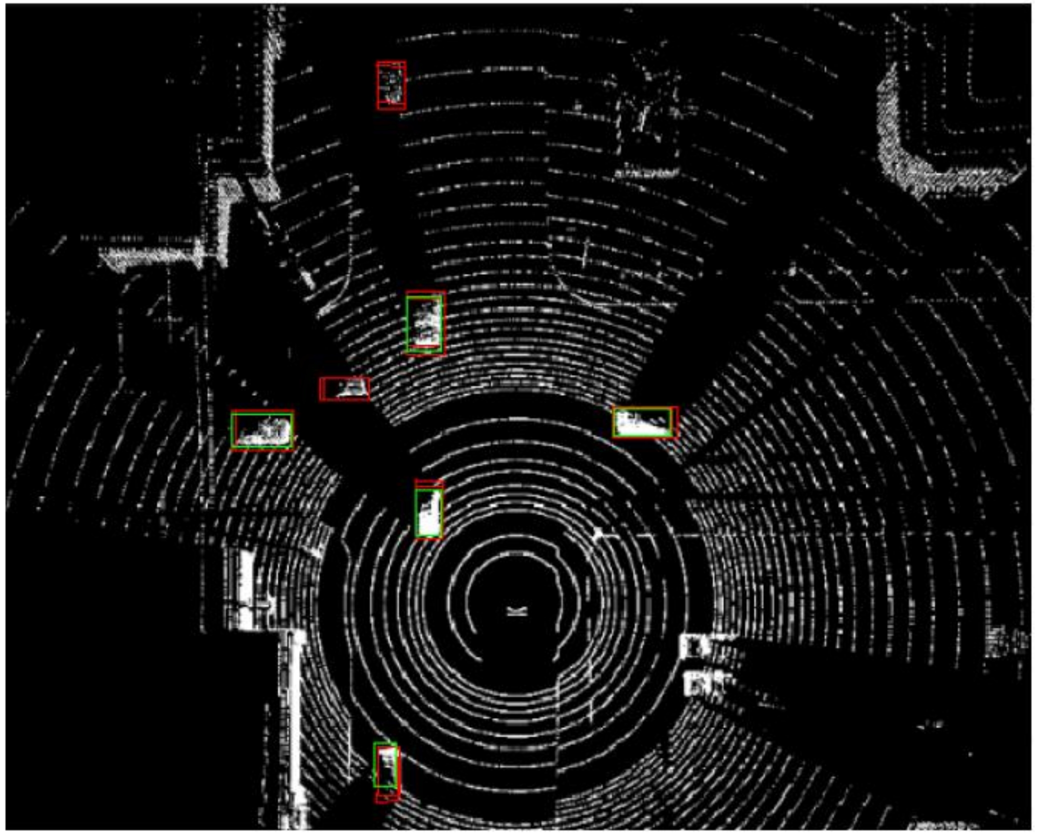
SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception

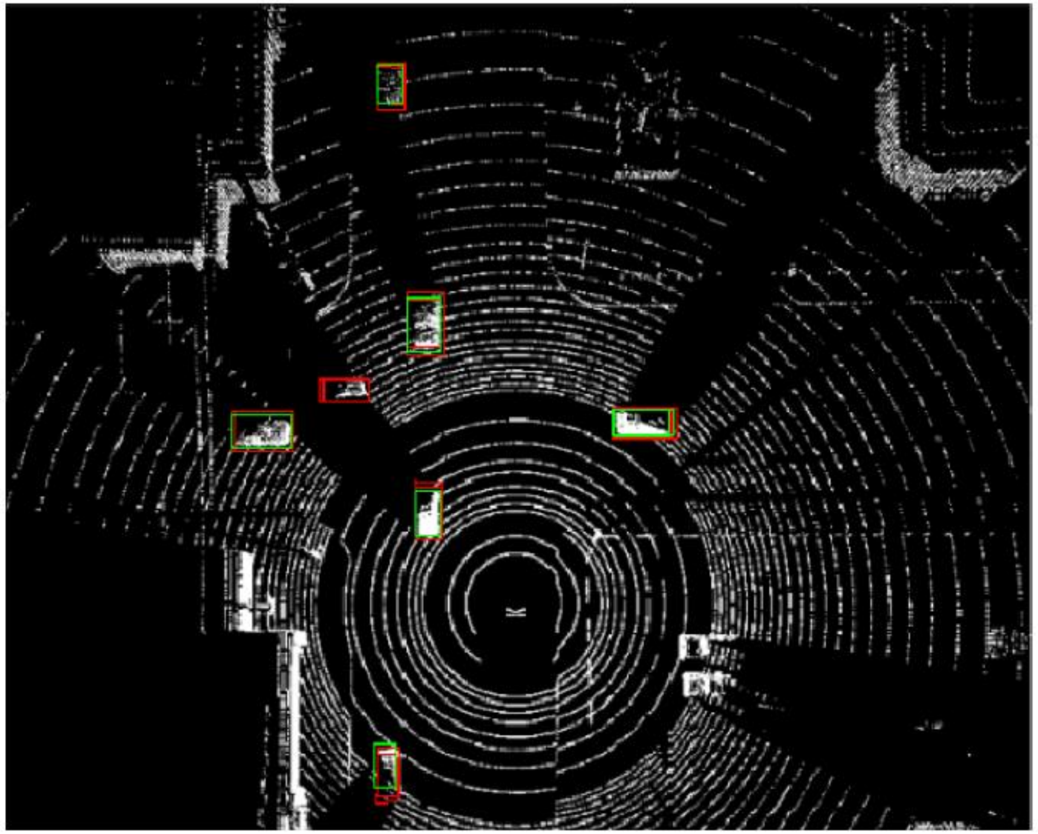
- Resultados

Dataset	PCSim LiDAR	DA	IoU	mAP@0.7
KITTI	HDL-64	-	↓4.29%	↓12.34%
	HDL-64	Real2Sim	**↑1.42%**	**↑10.26%**
NuScenes	HDL-32	-	↓8.33%	↓13.23%
	HDL-32	Real2Sim	↓8.33%	↓13.23%
	CARLA-32	-	↓9.52%	↓20.58%
	CARLA-32	Real2Sim	↓9.52%	↓20.58%
Pandaset	Pandar64	-	↓2.27%	↓8.45%
	Pandar64	Real2Sim	**0.0%**	**↑2.81%**

Funded by the European Union

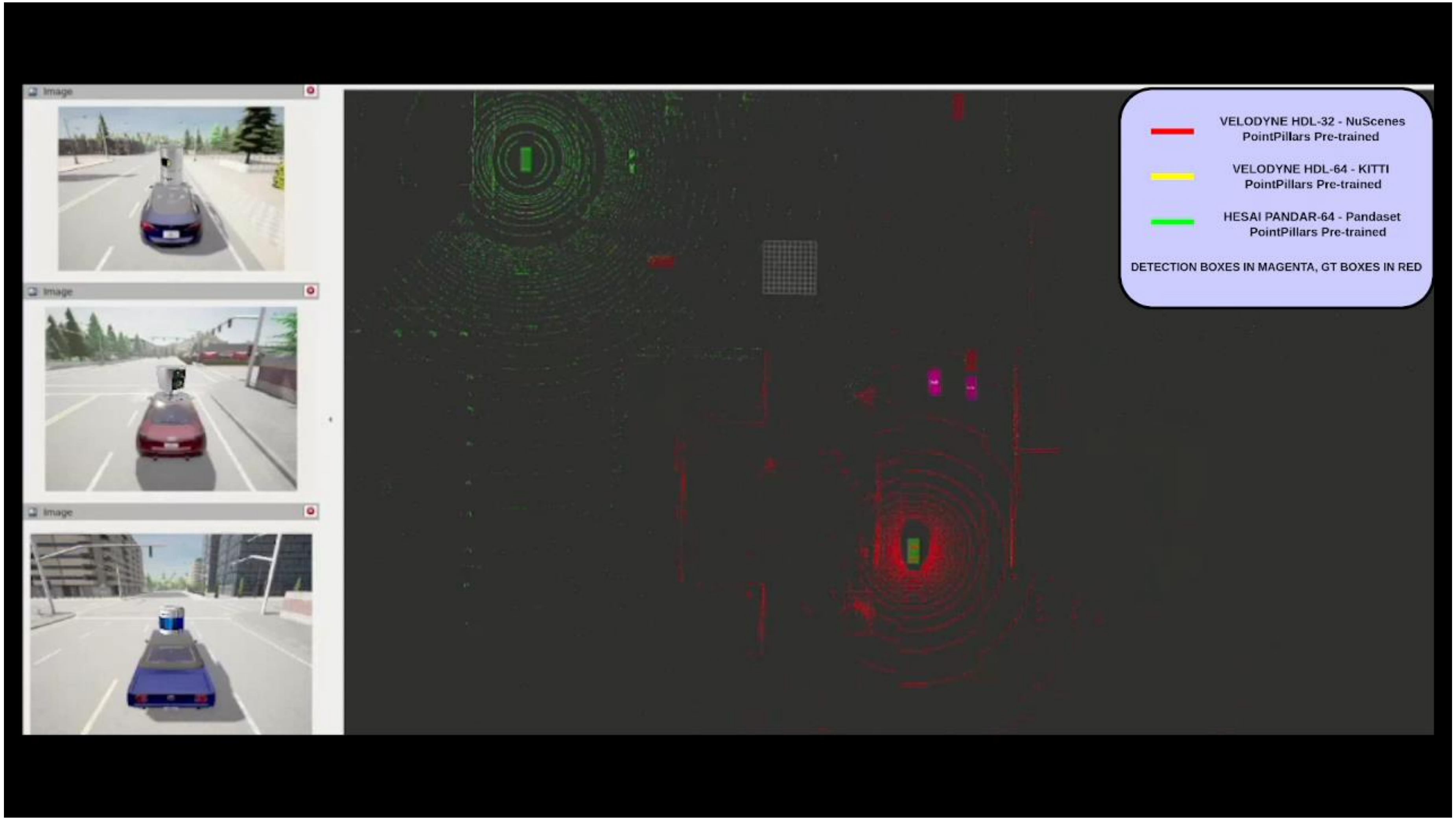
SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception





Funded by the European Union

SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception



Funded by the European Union

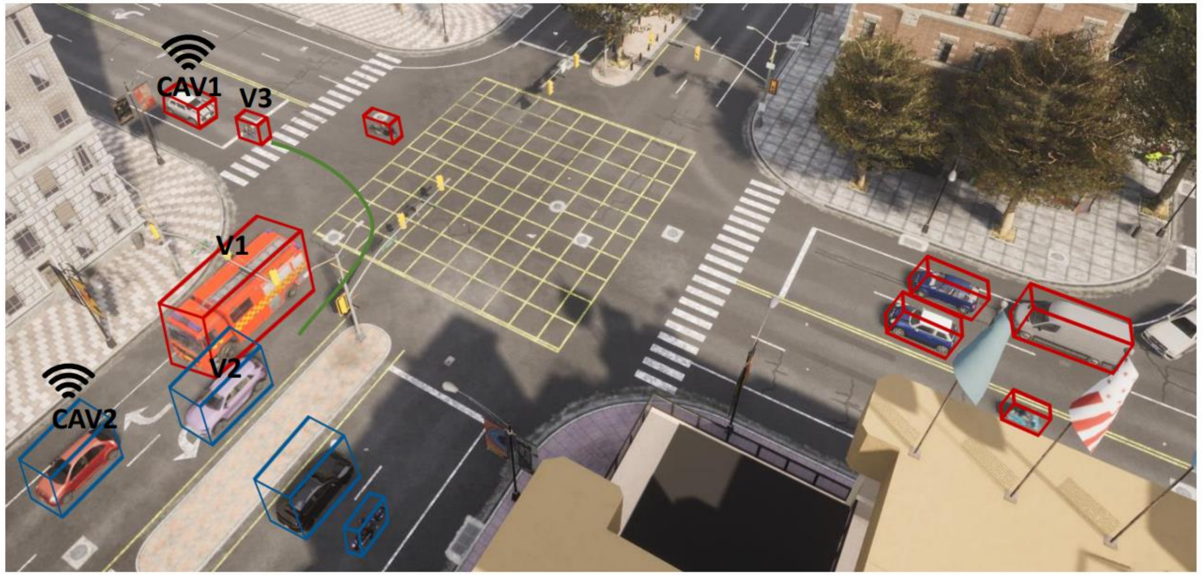
SimBusters: Bridging Simulation Gaps in Intelligent Vehicles Perception

- Conclusiones
  - Reducción del *gap* entre simulación y real
  - Método de *Domain Adaptation*
  - Futuro necesario para poder realizar más simulaciones

Funded by the European Union

Motion Prediction colaborativo

- Problemas de oclusiones
- Información más lejana
- Más contexto social



- Javier Araluce, Alberto Justo, Asier Arizala, Leonardo Gonzalez and Sergio Díaz, “Enhancing Motion Prediction by a Cooperative Framework”, in 2024 IEEE Intelligent Vehicles Symposium (IV), IEEE, June 2024.

Funded by the European Union

Enhanced Motion Prediction by Cooperative Framework

- Fusión de detecciones
  - Distancia euclídea

$$D_e = \sqrt{(x_{det} - x_{cav})^2 + (y_{det} - y_{cav})^2}$$

$$det = argmin(D_e^A, D_e^B)$$

  - *Bounding box clustering*

$$P = \{p_i\} \subset 3D_{box}^A \quad \forall i \in k < |\mathbb{Z}|$$

$$Q = \{q_j\} \subset 3D_{box}^B \quad \forall j \in l < |\mathbb{Z}|$$

$$det = max(\{dim(P), dim(Q)\})$$

Funded by the European Union

Enhanced Motion Prediction by Cooperative Framework

TABLE I: Comparison between Cooperative Perception-related Datasets.

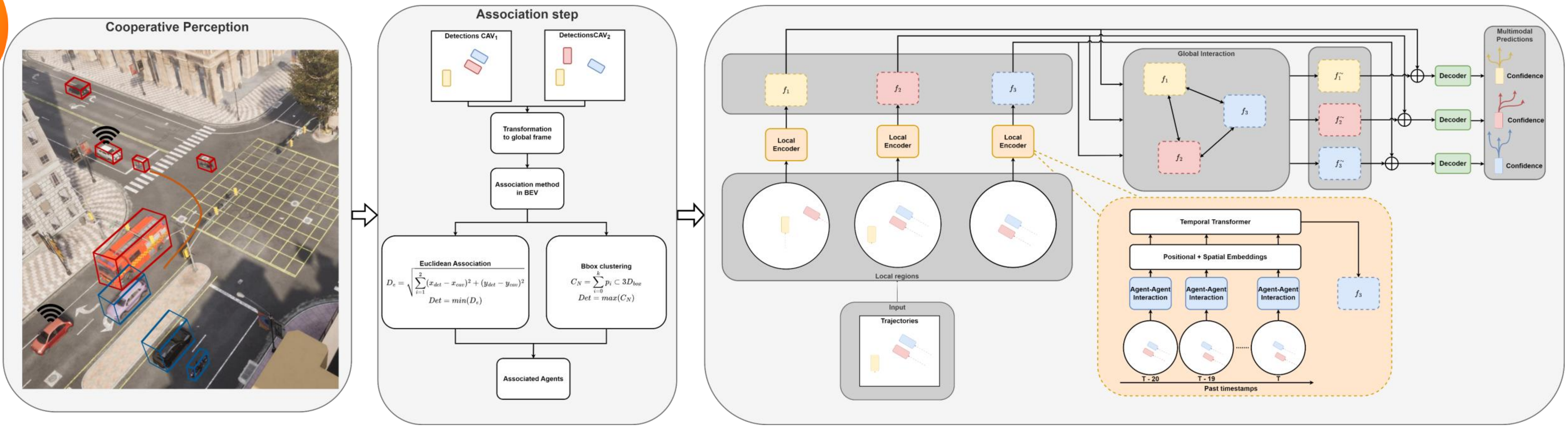
Dataset	Real/Sim	V2X	Size (km)	Lidar pcds	Maps	3D boxes	Classes	Locations
OPV2V [19]	Sim	V2V	-	11k	Yes	230k	1	CARLA
V2X-Sim [20]	Sim	V2V&I	-	10k	Yes	26.6k	1	CARLA
V2XSet [21]	Sim	V2V&I	-	11k	Yes	230k	1	CARLA
A9 Intersection [23]	Real	V2I	-	4.8k	No	57.4k	10	Hanover, Germany
DAIR-V2X [24]	Real	V2I	20	39k	No	464k	10	Beijing, CN
V2X-Seq [25]	Real	V2V&I	-	210k (seq)	No	20,301k (2D)	8	Beijing, CN
V2V4Real [15]	Real	V2V	410	20k	Yes*	240k	5	Ohio, USA

Notes: * indicates that the map are listed as public but they have not been released by the day of this work.

Funded by the European Union

Enhanced Motion Prediction by Cooperative Framework

EVENTS



Funded by the European Union

Enhanced Motion Prediction by Cooperative Framework

TABLE III: Comparison of methods on the V2V4Real dataset. We show the CAVs, the association method, the viewpoint, the number of actors considered and the performance metrics. The ”-” denotes that there is no association method used.

CAVs	Association	Viewpoint	Number Actors	minADE (m) ↓	minFDE (m) ↓	MR ↓	p-minADE ↓	p-minFDE ↓	p-MR ↓	brier-minADE (m) ↓	brier-minFDE (m) ↓
Tesla	-	Tesla	7.74	**1.14**	**2.22**	**0.32**	**2.84**	**3.91**	**0.87**	**1.80**	**2.88**
Astuff	-	Astuff	8.45	1.22	2.30	0.33	2.90	3.99	**0.87**	1.88	2.96
V2V	-	Tesla	14.58	1.34	2.51	0.33	3.02	4.19	**0.87**	2.00	3.17
V2V	Euclidean	Tesla	10.19	1.26	2.37	**0.32**	2.95	4.05	**0.87**	1.92	3.02
V2V	Bbox clustering	Tesla	10.19	1.26	2.37	**0.32**	2.95	4.06	**0.87**	1.92	3.03
V2V	-	Astuff	14.58	1.34	2.52	0.33	3.03	4.21	**0.87**	2.00	3.18
V2V	Euclidean	Astuff	10.19	1.27	2.38	**0.32**	2.95	4.07	**0.87**	1.92	3.04
V2V	Bbox clustering	Astuff	10.19	1.27	2.39	0.33	2.96	4.08	**0.87**	1.93	3.05

TABLE IV: Comparison of methods on the V2V4Real dataset normalised by the number of actors in the scene. We show the CAVs, the association method, the viewpoint and the performance metrics. The ”-” denotes that there is no association method used.

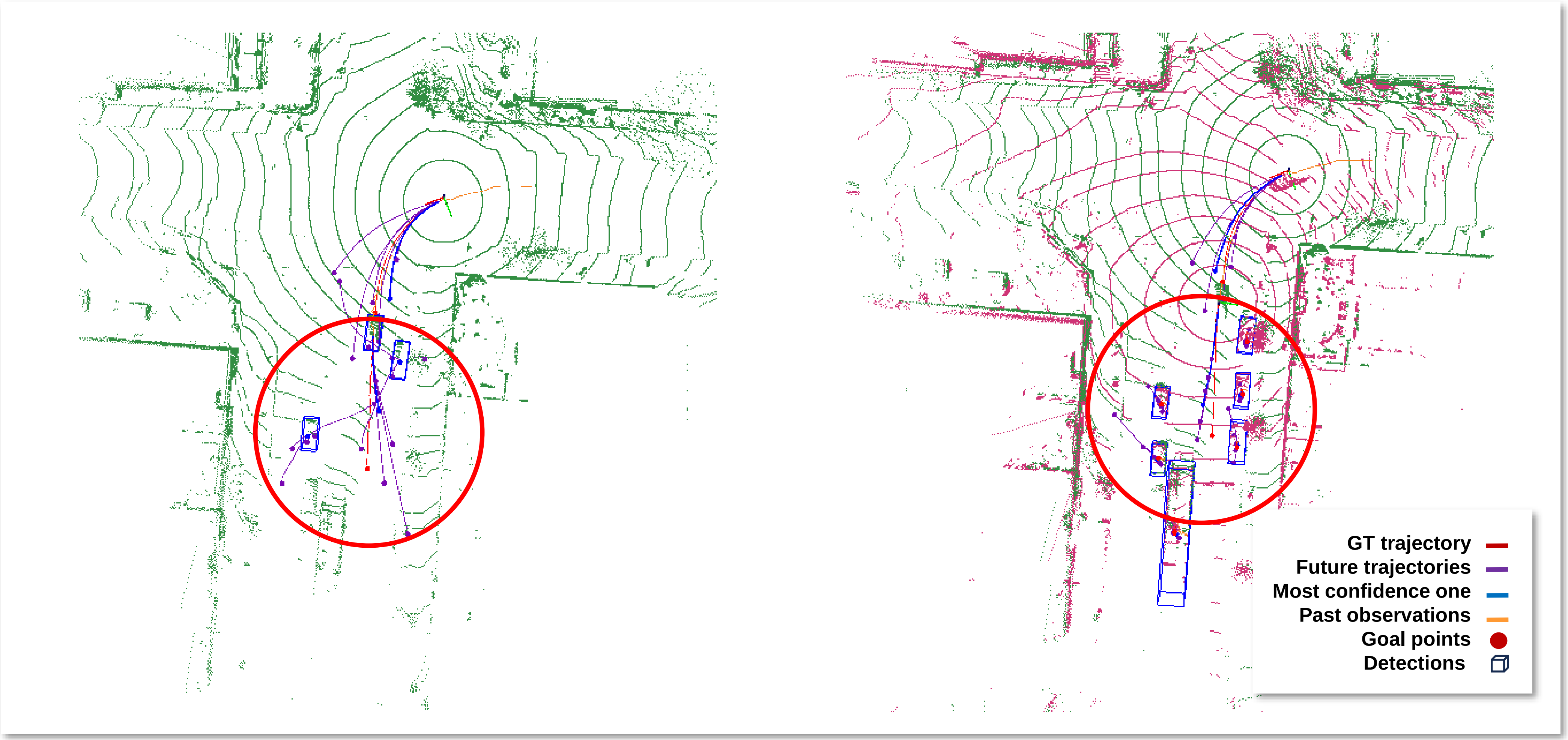
CAVs	Association	Viewpoint	minADE ↓	minFDE ↓	MR ↓	p-minADE ↓	p-minFDE ↓	p-MR ↓	brier-minADE ↓	brier-minFDE ↓
Tesla	-	Tesla	0.15	0.29	0.04	0.37	0.51	0.11	0.23	0.37
Astuff	-	Astuff	0.14	0.27	0.04	0.34	0.47	0.10	0.22	0.35
V2V	Euclidean	Tesla	**0.12**	**0.23**	**0.03**	**0.29**	**0.40**	**0.09**	**0.19**	**0.30**
V2V	Bbox clustering	Tesla	**0.12**	**0.23**	**0.03**	**0.29**	**0.40**	**0.09**	**0.19**	**0.30**
V2V	Euclidean	Astuff	0.13	0.24	**0.03**	**0.29**	**0.40**	**0.09**	**0.19**	**0.30**
V2V	Bbox clustering	Astuff	**0.12**	**0.23**	**0.03**	**0.29**	**0.40**	**0.09**	**0.19**	**0.30**

TABLE V: Increased performance through our V2V framework compared to a single vehicle, taking into account the number of vehicles. We show the comparison, the association method, the main CAV and the performance metrics.

Comparison	Association	Viewpoint	minADE	minFDE	MR	p-minADE	p-minFDE	p-MR	brier-minADE	brier-minFDE
V2V vs Tesla	Euclidean	Tesla	**16%**	**19%**	**22%**	**21%**	**21%**	**24%**	**19%**	**20%**
V2V vs Tesla	Bbox clustering	Tesla	**16%**	**19%**	**22%**	**21%**	**21%**	**24%**	**19%**	**20%**
V2V vs Astuff	Euclidean	Astuff	13%	14%	19%	15%	15%	17%	14%	14%
V2V vs Astuff	Bbox clustering	Astuff	14%	14%	19%	16%	15%	17%	15%	15%

Funded by the European Union

Qualitative results



Funded by the European Union

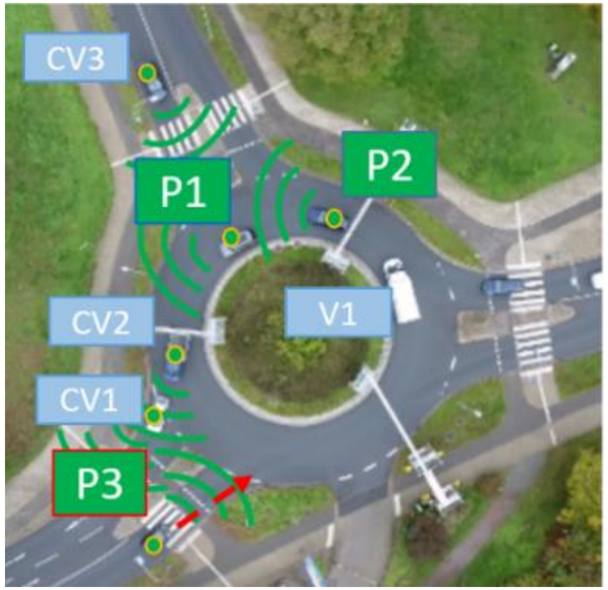
Enhanced Motion Prediction by Cooperative Framework

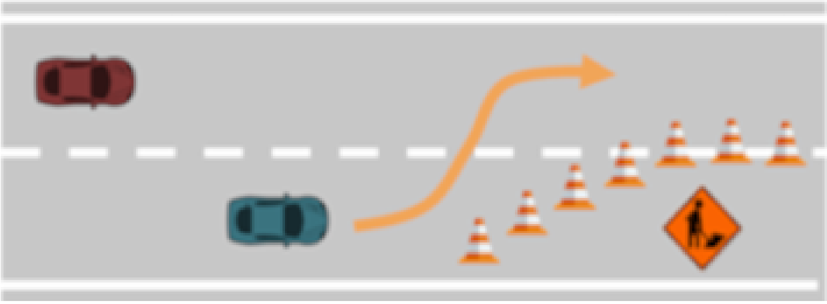
- Conclusiones
  - HiVT en un ambiente colaborativo
  - Dos métodos de asociación
  - V2V ayuda en Motion Prediction

Funded by the European Union

Experimentos/Demos

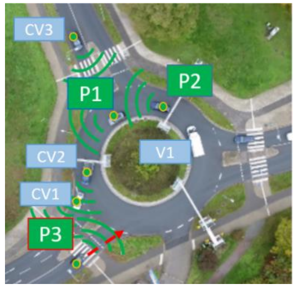
- Tres experimentos:
  - EXP2: rotonda con un platoon roto
  - EXP4: zona de obras
  - EXP5: negociar la entrada en una intersección





Funded by the European Union

EXP2

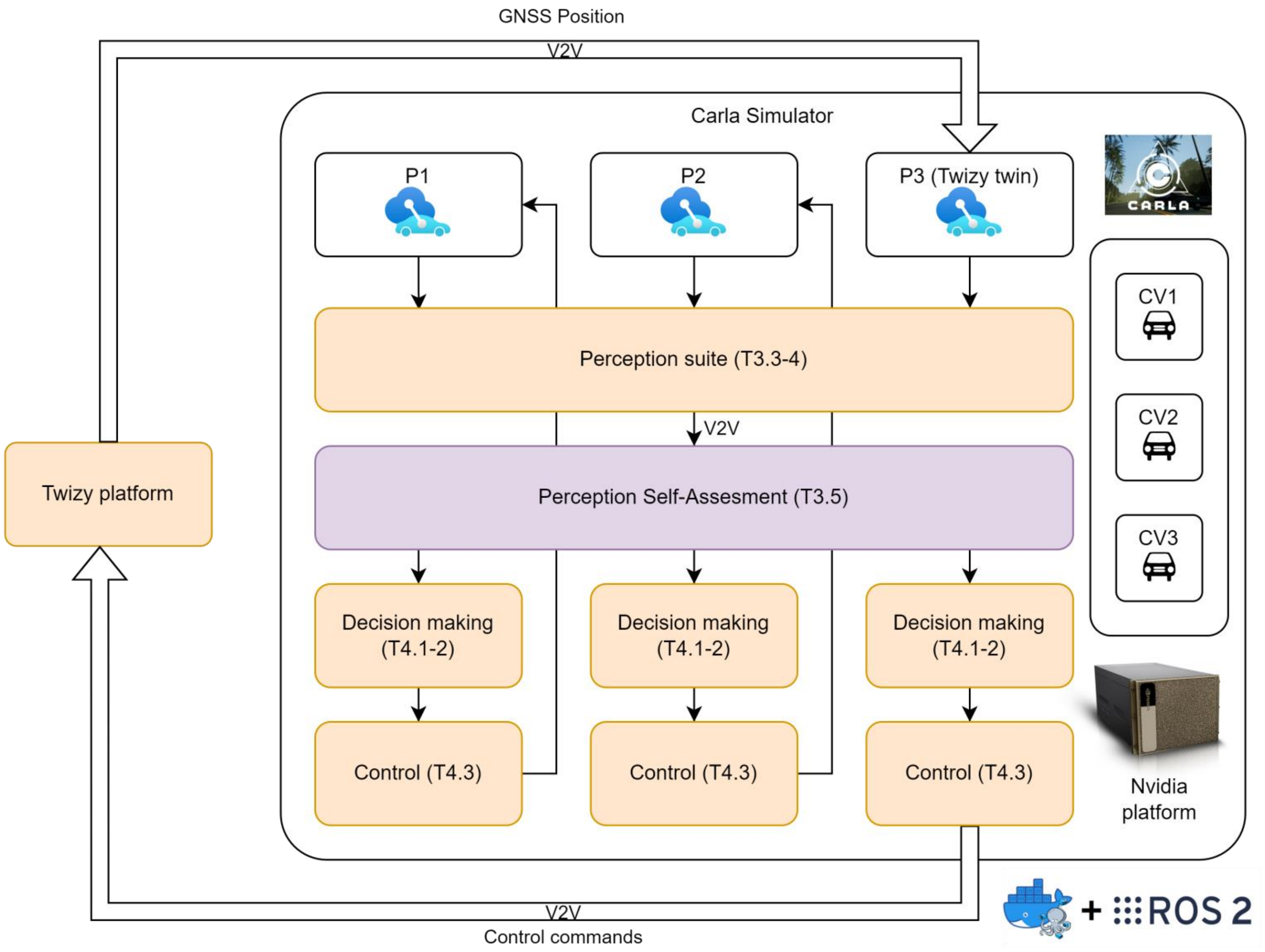


- P1, P2 & CV: Simulados en Carla

- P3: Renault Twizy

Funded by the European Union

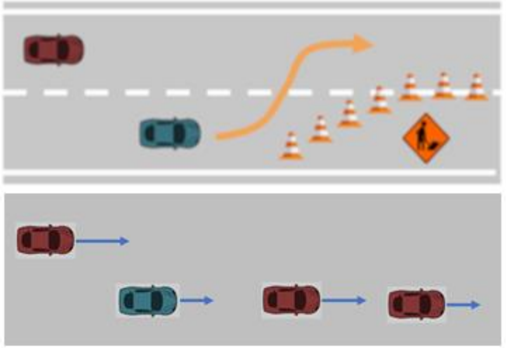
EXP2



Funded by the European Union

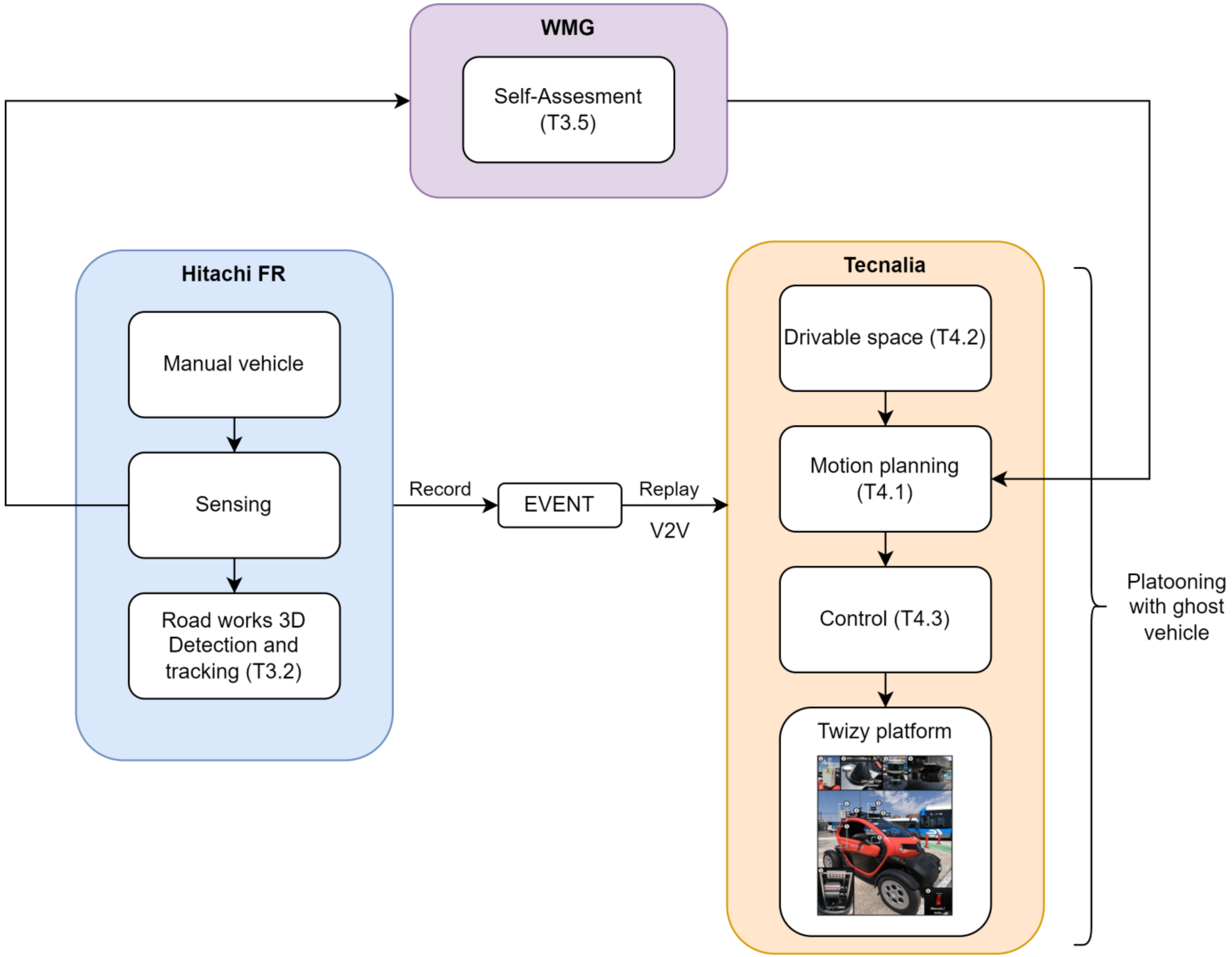
EXP4

- Un coche en manual detecta una zona de obras
  - Nosotros lo replicamos en un ambiente controlado utilizando sus detecciones



Funded by the European Union

EXP4



Funded by the European Union

www.events-project.eu

EVENTSproject22

@EVENTSproject22

EVENTS project

EVENTS

Thank you for your attention!

tecnalia

MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

Javier Araluce, Tecnalia,
javier.araluce@tecnalia.com

Funded by the European Union

This project has received funding under grant agreement No 101069614. It is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.