Anteproyecto de Trabajo Fin de Máster

Santiago Montiel Marín

1 de diciembre de 2022

Título: Aprendizaje profundo para detección de objetos 3D mediante fusión de radar de automoción y cámara

Título en inglés: Deep Learning based 3D Object Detection for Automotive Radar and Camera Fusion

Departamento: Departamento de Electrónica

Autor: Santiago Montiel Marín

Tutores: Luis Miguel Bergasa Pascual y Ángel Llamazares Llamazares

1 Introducción

Los radares de longitud de onda milimétrica (*mmWave RADAR*, en inglés) son ampliamente utilizados en la industria automotriz para el desarrollo de sistemas avanzados de asistencia a la conducción (ADAS). Estos sistemas, entre los que se incluyen el frenado automático de emergencia (AEB) o el control de velocidad de crucero adaptativa (ACC), hacen uso intensivo del radar para la percepción del entorno y están presentes en multitud de vehículos producidos en masa. [1]

Por otro lado, las tecnologías de conducción autónoma han experimentado un desarrollo importante durante los últimos años, gracias a los esfuerzos de la industria y la academia y del crecimiento del aprendizaje profundo (DL). En este nuevo paradigma, la mayor parte de las soluciones de percepción del entorno se han basado en el uso de sensores LiDAR y cámaras, o una fusión de ambos, siendo el radar relegado a un segundo plano.

Esto se debe a que los radares de uso comercial presentes en el mercado proporcionan como información una nube de puntos tridimensional (rango, azimuth y velocidad Doppler) de densidad dispersa, puesto que una de las debilidades de este sensor es la escasa resolución angular. A priori, esta información no es lo suficientemente rica para obtener una comprensión profunda del entorno que rodea al vehículo y que este realice comportamientos complejos de forma autónoma.

Para afrontar esta situación, el sector del radar de automoción ha conseguido avances recientes y es posible que la década de 2020 sea la década en la que el aprendizaje profundo impulse la percepción en vehículos autónomos. Esta afirmación se respalda observando las distintas líneas de trabajo que aparecen en la literatura reciente.

- El surgimiento de una nueva generación de sensores radar 4D para automoción. La industria está en proceso de desarrollar nuevos sensores que mantienen las ventajas tradicionales del radar, como la inferencia de velocidad mediante efecto Doppler y la resolución en rango, y tienen mejoras significativas en cuanto a las desventajas que lo relegaron, como la resolución en azimuth y la aparición de elevación. Con estas mejoras, los radares proporcionan nubes de puntos 4D de alta resolución.

- El creciente interés de la academia en el sector. Desde 2019, el número de artículos académicos de Deep Learning para percepción con radar ha ido en aumento en las conferencias de mayor prestigio en las temáticas de transporte inteligente (como IV, ITSC), robótica (IROS, ICRA) y visión artificial (como CVPR, ICCV, ECCV). Además, estos artículos aplican enfoques novedosos en cuanto al tratamiento de los datos radar y las técnicas Deep Learning maduras procedentes de sectores como la visión artificial para mejorar la percepción del entorno.

- La aparición de nuevos trabajos que implementan la fusión sensorial de radar de automoción y cámara monocular para las tareas de comprensión de la escena mediante el uso de técnicas Deep Learning. Mientras que el radar aporta información sobre la posición en el espacio 3D y su velocidad, la cámara aporta características como la apariencia y la relación de aspecto en el plano imagen.

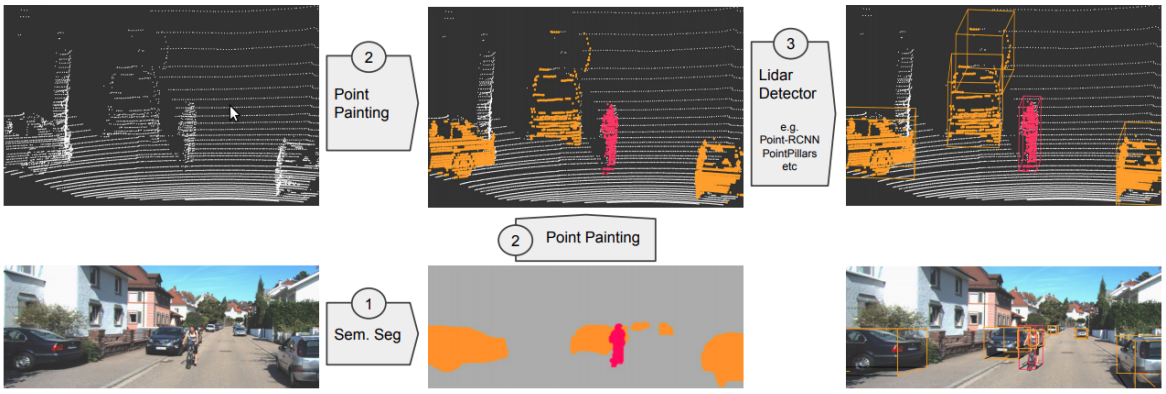

Figura 2: **Fusion sensorial** de imagen de cámara y nube de puntos de lidar. Adaptado de [2].

2 Objetivos y desarrollo

El objetivo fundamental de este trabajo es el estudio, implementación y evaluación de un sistema Deep Learning con múltiples redes neuronales (multi-etapa) para la detección de objetos 3D en el espacio tridimensional a partir de los datos de nubes de puntos 3+1D de radar, comerciales y de nueva generación, y de las imágenes de una cámara monocular. Este enfoque se alinea con las tendencias de la literatura reciente [3] [4] [5] [6], que buscan diseñar métodos para que la fusión sensorial de radar y cámara monocular tenga un rendimiento similar al que tienen los métodos basados en lidar. Por tanto, el propósito final del sistema es la detección y clasificación

de objetos estáticos y dinámicos que rodean al vehículo mediante Deep Learning con estos dos sensores. Un resultado de trabajos similares, con la diferencia de tomar la nube de puntos de un lidar, puede verse en la Figura 2.

Con mayor detalle, se procede a la formulación de los objetivos específicos del proyecto:

- Realizar un estudio de la fusión sensorial entre nube de puntos y cámara monocular mediante proyecciones geométricas.
- Diseñar e implementar un *framework* de entrenamiento de modelos de detección de objetos 3D en *PyTorch* que permita experimentar con diferentes técnicas de entrenamiento, arquitecturas del estado del arte, funciones de pérdidas y optimizadores.
- Preentrenar redes neuronales para cámara monocular en tareas que permitan colorear posteriormente la nube de puntos de radar en datasets de referencia.
- Evaluar los modelos previamente entrenados en datasets de nubes de puntos radar utilizando métricas procedentes del estado del arte de la visión artificial.
- Mejorar los resultados mediante la experimentación, implementando herramientas de análisis del error.
- Formular y llevar a cabo experimentos que permitan una comprensión más profunda del enfoque propuesto comparando diferentes arquitecturas y datasets.
- Realizar una comparación con otros métodos establecidos en el estado del arte, ya sean tradicionales o basados en Deep Learning con otro enfoque.

3 Metodología y plan de trabajo

El proyecto a realizar tiene una duración prevista de 8 meses: desde diciembre de 2022 hasta julio de 2023, ambos meses inclusive. Para organizar su desarrollo, este se divide en fases que son alineadas con los objetivos descritos en la sección 2.

1. Formación inicial en la temática. (1 mes)
   - Introducción al *framework* de aprendizaje automático *PyTorch*.
   - Consulta bibliográfica del estado del arte de detección de objetos mediante Deep Learning para radar y cámara monocular.
2. Estudio de las bases de datos descritas en la sección 4 y adaptación de los datos desde su formato original al deseado. (1 mes)
3. Desarrollo de un *framework* de detección de objetos 3D en *PyTorch*. (2 meses)
   - Implementación de los bucles de entrenamiento, funciones de pérdida y optimizadores.
   - Implementación de las arquitecturas de cámara monocular, como *YOLOv5*, y detección de objetos del arte, como *PointPillars*.

- Desarrollo de *scripts* de validación e inferencia o predicción ante nuevos datos.

4. Evaluación de los modelos mediante la implementación de algoritmos de detección de objetos 2D. (1 mes)

5. Implementación de una herramienta de análisis de error para contribuir a una mayor explicabilidad sobre el comportamiento de los modelos. (1 mes)

6. Implementación de una solución clásica o basada en Deep Learning con otro enfoque, a definir todavía, para comparar con el desempeño del enfoque propuesto. (1 mes)

7. Documentación del desarrollo del proyecto, los estudios y experimentos realizados y resultados obtenidos en la memoria del Trabajo Final de Máster. (1 mes)

4 Medios

Para la correcta realización del trabajo proyectado se necesitará utilizar un conjunto de medios y herramientas. Por un lado, las herramientas hardware que van a ser necesarias para desarrollar este proyecto son las siguientes:

- Ordenador personal: memoria RAM de 32GB, tarjeta gráfica NVIDIA RTX 3090Ti.
- Plataforma de conducción autónoma del grupo de investigación *RobeSafe*.
- Radares de automoción comerciales o experimentales, si es posible.
- Cámara monocular *ZED* para robótica y vehículos autónomos.

Por otro lado, se necesitarán los siguientes recursos software:

- Sistema operativo de código abierto *Linux*.
- Editor de código fuente *Visual Studio Code*.
- Software de control de versiones *Git*.
- Lenguaje de programación *Python*.
- Framework de aprendizaje automático *PyTorch*.
- Biblioteca para manipulación y análisis de datos *Pandas*.

Para poder realizar el entrenamiento de los algoritmos de aprendizaje profundo, se hará uso de los siguientes *datasets* o bases de datos, cuyo acceso puede ser público o privado mediante autorización de los autores:

- **Cityscapes** [7] (2016): este dataset es el referente en cuanto a la tarea de segmentación semántica se refiere. Es un conjunto de datos de gran escala que contiene imágenes tomadas desde el frontal de un vehículo en diferentes ciudades europeas, presentadas en formato vídeo. Tiene anotaciones a nivel de píxel, instancia y panóptica, permitiendo el entrenamiento de redes en estos tres tipos de tareas de segmentación.

- **View of Delft** [8] (2022): elaborado por la Universidad Técnica de Delft. Cuenta con datos de un radar *ZF Gen13 Short Range*, cuya información está sincronizada con un lidar, cámara, odometría y GPS. Tiene anotaciones 2D para la tarea de detección de objetos en el plano imagen, y 3D para la tarea detección de objetos en el espacio tridimensional. Se compone de unas 7000 escenas y se centra en la clasificación de *Vulnerable Road Users (VRU)* en entornos urbanos.

- **TJ4DRadSet** [9] (2022): elaborado por la Universidad de Tongji, cuenta con datos de un radar *Oculii Eagle Long Range*, sincronizados con cámara y lidar. Los objetos de interés están anotados mediante *Bounding Boxes* 3D y *track-ids* que permiten realizar seguimiento de objetos. La utilización de este dataset queda supeditado al lanzamiento del mismo, puesto que en el momento de la escritura de este documento, el *dataset* no ha sido liberado.

- **Conjunto de datos Huawei-UAH** (2022): una serie de escenarios capturados por radares experimentales de la compañía Huawei en el marco de cooperación con la Universidad de Alcalá, grabados en las instalaciones del Campus Exterior de esta última.

Referencias

[1] Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue, "Towards deep radar perception for autonomous driving: Datasets, methods, and challenges," *Sensors*, vol. 22, no. 11, p. 4208, 2022.

[2] S. Vora, A. H. Lang, B. Helou, and O. Beijbom, "Pointpainting: Sequential fusion for 3d object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4604–4612.

[3] N. Scheiner, F. Kraus, N. Appenrodt, J. Dickmann, and B. Sick, "Object detection for automotive radar point clouds–a comparison," *AI Perspectives*, vol. 3, no. 1, pp. 1–23, 2021.

[4] F. Nobis, M. Geisslinger, M. Weber, J. Betz, and M. Lienkamp, "A deep learning-based radar and camera sensor fusion architecture for object detection," in *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2019, pp. 1–7.

[5] R. Nabati and H. Qi, "Centerfusion: Center-based radar and camera fusion for 3d object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1527–1536.

[6] S. Baratam, "Radar-guided monocular depth estimation and point cloud fusion for 3d object detection," 2022.

[7] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[8] A. Palffy, E. Pool, S. Baratam, J. F. Kooij, and D. M. Gavrila, "Multi-class road user detection with 3+ 1d radar in the view-of-delft dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4961–4968, 2022.

[9] L. Zheng, Z. Ma, X. Zhu, B. Tan, S. Li, K. Long, W. Sun, S. Chen, L. Zhang, M. Wan *et al.*, "Tj4dradset: A 4d radar dataset for autonomous driving," *arXiv preprint arXiv:2204.13483*, 2022.