

Optimización de Sistemas Embebidos en Vehículos Autónomos: Estrategias de Co-Diseño y Tecnologías Avanzadas

Marbin Javier Arévalo Guerrero
Escuela de Ingeniería de Sistemas
Universidad Industrial de Santander
Bucaramanga, Colombia

Abstract—Este artículo realiza una revisión exhaustiva de los vehículos autónomos, destacando las tecnologías clave y los desafíos asociados. Se enfoca en los sistemas embebidos y examina estrategias de co-diseño, como la integración de FPGA y enfoques CPU/FPGA, para mejorar el rendimiento. Se analizan investigaciones del MIT (Instituto Tecnológico de Massachusetts) y estrategias específicas de Collin et al., así como la solución HALP de NVIDIA. El objetivo principal es identificar problemas y explorar cómo diversos autores y desarrolladores han abordado estos desafíos, contribuyendo al conocimiento y desarrollo de soluciones efectivas para vehículos autónomos.

Index Terms—FPGA (Field-Programmable Gate Arrays), NN (Redes Neuronales), CNN (Redes Neuronales Convolucionales), RNN (Redes Neuronales Recurrentes), HALP (Hardware-Aware Latency Purning), Deep Learning (Aprendizaje Profundo), Co-Diseño HW/SW (Co-Diseño de Hardware y Software)

I. INTRODUCCIÓN

La era de la conducción autónoma ha irrumpido en la industria automotriz, desencadenando una revolución en la forma en que concebimos y experimentamos el desplazamiento vehicular. Este cambio de paradigma, sin embargo, no está exento de desafíos intrincados, especialmente en lo que respecta a los sistemas embebidos que constituyen la columna vertebral de la autonomía vehicular.

En este contexto, nuestro artículo se sumerge en los fundamentos de los vehículos autónomos, destacando la complejidad de los problemas que surgen en la intersección de la inteligencia artificial, la ingeniería de sistemas y la computación embebida. Nos centramos en el papel crucial de los sistemas embebidos, también conocidos como Sistemas Ciber-Físicos (CPS), delineando su función en el reconocimiento espacial, un componente esencial para la conducción autónoma.

Exploramos a fondo los desafíos clave que enfrentan estos sistemas embebidos, desde la limitación de recursos computacionales hasta las estrictas restricciones de seguridad, y cómo estos desafíos afectan la velocidad, eficiencia y costos de los vehículos autónomos. Para abordar estos problemas, nos sumergimos en estrategias de co-diseño, centrándonos especialmente en la integración de dispositivos FPGA y en enfoques CPU/FPGA para optimizar tareas críticas.

A lo largo del artículo, destacamos investigaciones clave, como el análisis detallado realizado por el MIT sobre arquitecturas de hardware y software en sistemas de conducción

autónoma, así como las estrategias específicas propuestas por Collin et al., que ofrecen soluciones tangibles a los desafíos identificados.

Además, exploramos la esfera del Deep Learning y las Redes Neuronales, fundamentales para la capacidad de los vehículos autónomos para procesar información de manera eficiente. Presentamos la solución HALP de NVIDIA como un ejemplo destacado de cómo las estrategias de co-diseño pueden optimizar el rendimiento de estas tecnologías avanzadas en sistemas embebidos, superando las limitaciones de recursos.

En última instancia, este artículo se propone como una guía comprensiva, revelando las complejidades y ofreciendo soluciones innovadoras para el diseño y desarrollo de sistemas embebidos en vehículos autónomos. Desde la exploración de estrategias específicas hasta la optimización de análisis espacial mediante Deep Learning, delineamos un camino hacia vehículos autónomos más eficientes, seguros y capaces de enfrentar los desafíos del futuro de la movilidad.

II. SISTEMA EMBEBIDO Y DESAFÍOS EN VEHÍCULOS AUTÓNOMOS

En esta sección, exploraremos los fundamentos de los sistemas embebidos y su papel crucial en el contexto de los vehículos autónomos, también conocidos como Sistemas Ciber-Físicos (CPS, por sus siglas en inglés). Un sistema embebido, en este contexto, se refiere a un sistema computacional integrado en el vehículo, diseñado para cumplir funciones específicas sin requerir intervención humana directa. El reconocimiento espacial es esencial para la conducción autónoma, y los sistemas embebidos desempeñan un papel vital al procesar información de sensores como RADAR, LiDAR (que utiliza láseres para obtener información de distancia y profundidad) y cámaras. Estos sensores permiten la creación de una representación precisa del entorno del vehículo en tiempo real, ofreciendo una visión de 360 grados. Sin embargo, este proceso enfrenta desafíos significativos, como la capacidad de cómputo limitada, restricciones de memoria y consumo eléctrico, factores que impactan directamente la velocidad y eficiencia del sistema. Uno de los problemas recurrentes es la velocidad de procesamiento, y este desafío es acentuado por la necesidad de toma de decisiones en

tiempo real en entornos dinámicos. Gracias a una perspectiva detallada proporcionada por diversos autores sobre cómo estos desafíos impactan el rendimiento de los sistemas embebidos en vehículos autónomos, se destaca la necesidad de optimizar la latencia y los costos sin comprometer la seguridad, abordando así una de las preocupaciones fundamentales en este dominio. El MIT en uno de sus artículos publicados [5], ofrece una exploración exhaustiva de las arquitecturas de hardware y software utilizadas en sistemas de conducción autónoma, así mismo nos muestra un análisis donde finalmente se concluye que uno de los problemas más grandes de esos sistemas es la latencia y los costos bajo estrictas restricciones de seguridad, el estudio presenta resultados cuantificables que revelan mejoras sustanciales en términos de velocidad de procesamiento y eficiencia, marcando un umbral en cuanto a qué tan rápido debe ser el tiempo de reacción y por qué. En la siguiente sección, nos sumergiremos más profundamente en algunas estrategias que pueden ser usadas para abordar estos desafíos, brindando así una comprensión detallada de cómo las soluciones propuestas pueden influir positivamente en el rendimiento y fiabilidad de los vehículos autónomos.

III. ESTRATEGIAS PARA OPTIMIZAR SISTEMAS EMBEBIDOS EN VEHÍCULOS AUTÓNOMOS

En esta sección, nos tomaremos las estrategias fundamentales propuestas por *C Collin, Anne et al.* en el artículo [5]. Este análisis permitirá entender en detalle la importancia de los FPGA, explorar estrategias CPU/FPGA y su papel en la optimización de tareas críticas para vehículos autónomos.

FPGA y Estrategias CPU/FPGA

Los FPGA (Field-Programmable Gate Arrays) son dispositivos de hardware flexibles que permiten la programación de lógica digital después de la fabricación. En el contexto de los vehículos autónomos, la capacidad de reconfiguración de los FPGA se vuelve crucial. El artículo [3] profundiza en cómo la combinación de CPU y FPGA puede optimizar tareas específicas.

La estrategia CPU/FPGA aborda la necesidad de ejecutar algoritmos críticos para la conducción autónoma de manera eficiente. Mientras que la CPU maneja tareas generales, el FPGA puede acelerar funciones específicas como el procesamiento de imágenes, el reconocimiento de objetos y la toma de decisiones en tiempo real. Esto permite una distribución eficiente de la carga de trabajo. Sin embargo, como veremos más adelante, algunas soluciones presentan nuevos desafíos.

Estrategias Específicas para la Optimización de Sistemas Embebidos

El artículo [5] propone estrategias específicas para superar los desafíos de los sistemas embebidos en vehículos autónomos. Se destaca la importancia de optimizar la latencia y los costos, considerando las estrictas restricciones de seguridad. La investigación aborda la necesidad de tomar decisiones rápidas en entornos dinámicos. Entre las estrategias específicas se incluyen la paralelización de tareas críticas y la

implementación eficiente de algoritmos en el hardware. Estos enfoques buscan maximizar el rendimiento sin comprometer la seguridad, un aspecto crucial para la adopción exitosa de la conducción autónoma.

Desafíos y Soluciones en la Paralelización de Tareas Críticas y Aceleración 2D/3D con FPGA

La implementación de la paralelización en sistemas embebidos plantea desafíos específicos que requieren soluciones cuidadosas. En el contexto de los vehículos autónomos y los sistemas ciber-físicos (CPS), es fundamental abordar estos desafíos para garantizar un rendimiento eficiente y una toma de decisiones en tiempo real. El artículo, Cabanes [3] explora en profundidad una metodología de co-diseño de hardware y software que aborda estos desafíos de manera integral. En este enfoque, se destaca la necesidad de equilibrar la mejora de la eficiencia computacional con la gestión de la latencia, especialmente en tareas críticas para la conducción autónoma. Se propone un enfoque híbrido CPU/FPGA que aprovecha las ventajas de la paralelización sin comprometer la latencia en situaciones cruciales. La metodología se presenta como una herramienta valiosa para la prototipación eficiente de CPS, con aplicaciones directas en la optimización de tareas específicas.

En la misma línea, el artículo subraya el papel fundamental de los FPGA en la aceleración del procesamiento 2D y 3D, aspecto crucial para el reconocimiento espacial en la conducción autónoma. Los FPGA permiten una mejora significativa en la eficiencia del procesamiento, destacándose en escenarios donde la rapidez y precisión son esenciales. Ejemplos específicos, como el manejo de nubes de puntos LiDAR para la detección de objetos, ilustran cómo la combinación estratégica de CPU y FPGA puede revolucionar el procesamiento de datos complejos en entornos 2D y 3D.

En la siguiente sección, tomaremos en cuenta uno de los pilares en la navegación autónoma de vehículos: *El Deep Learning y las Redes Neuronales*. Indagaremos en su utilidad y en cómo se han desarrollado soluciones de co-diseño HW/SW con el propósito especializado de resolver problemas que presenta el uso de estas tecnologías.

IV. OPTIMIZACIÓN DEL ANÁLISIS ESPACIAL EN VEHÍCULOS AUTÓNOMOS MEDIANTE DEEP LEARNING

Deep Learning y Redes Neuronales en Vehículos Autónomos

La implementación exitosa de vehículos autónomos se basa en la capacidad de los sistemas embebidos para procesar información de manera eficiente y tomar decisiones precisas en tiempo real. En este contexto, el papel crucial del deep learning y las redes neuronales se convierte en un componente esencial para la conducción autónoma.

El deep learning, una rama de la inteligencia artificial (IA), se ha destacado como una tecnología clave en la mejora de la capacidad de las máquinas para aprender y tomar decisiones complejas. En el ámbito de los vehículos autónomos, la aplicación del deep learning implica la capacitación de modelos de inteligencia artificial para comprender y responder a patrones complejos en los datos sensoriales.

Las redes neuronales, inspiradas en el funcionamiento del cerebro humano, son los bloques fundamentales del deep learning. La capacidad de estas redes para aprender y adaptarse a través de capas de procesamiento las hace ideales para tareas como el reconocimiento de objetos, la clasificación de imágenes y la toma de decisiones autónomas.

En vehículos autónomos, las redes neuronales son esenciales para procesar datos visuales, como imágenes de cámaras y sensores, permitiendo al vehículo "ver" y entender su entorno. Estas redes aprenden automáticamente características jerárquicas, como bordes y formas, para realizar tareas como detección de objetos y reconocimiento de patrones. En la implementación específica para vehículos autónomos, se emplean varios tipos de redes neuronales para abordar tareas específicas:

- **Redes Neuronales Convolucionales (CNN):** Diseñadas para el procesamiento eficiente de datos espaciales, las CNN son esenciales para el reconocimiento de patrones en imágenes y videos. Su capacidad para capturar características relevantes las hace fundamentales en la detección de objetos a partir de datos visuales, como las imágenes capturadas por cámaras.
- **Redes Neuronales Recurrentes (RNN):** Utilizadas para tareas que implican secuencias de datos, las RNN son valiosas en la predicción y toma de decisiones basadas en la información temporal. En el contexto de vehículos autónomos, podrían aplicarse en situaciones que requieren un entendimiento del contexto temporal, como la anticipación de movimientos de peatones o vehículos.
- **Redes Neuronales Generativas (GAN):** Siendo capaces de generar datos nuevos a partir de patrones aprendidos, las GAN son útiles en la creación de conjuntos de datos sintéticos para el entrenamiento de modelos. Esto es crucial en entornos de conducción autónoma, donde la diversidad de escenarios y condiciones puede ser limitada.

Aunque el deep learning ofrece capacidades avanzadas, su integración en sistemas embebidos presenta desafíos únicos. La limitación de recursos computacionales, como potencia de procesamiento y memoria, requiere estrategias específicas para garantizar la eficiencia y la respuesta en tiempo real.

Optimización de Cálculos por NVIDIA: HALP

NVIDIA aborda eficientemente estos cálculos al aprovechar sus GPU especializadas, como las de la serie NVIDIA Drive, diseñadas específicamente para cargas de trabajo de conducción autónoma.

Estas GPU utilizan arquitecturas como CUDA para ejecutar operaciones de redes neuronales de manera altamente paralela, acelerando significativamente el procesamiento. Además, se implementan técnicas de precisión reducida, como FP16 (media precisión), para realizar cálculos con menor demanda computacional, sin sacrificar la precisión del modelo. Esto es especialmente crucial en sistemas embebidos donde los recursos son limitados.

HALP optimiza aún más el rendimiento al personalizar el diseño de la red y los algoritmos para adaptarse a las

características de las GPU de NVIDIA, asegurando una ejecución eficiente y rápida de las operaciones de CNN. Esta combinación de hardware especializado y estrategias de diseño específicas contribuye significativamente a la mejora del análisis espacial en vehículos autónomos.

En resumen, HALP no solo potencia la capacidad de los vehículos autónomos para comprender patrones complejos, sino que también aborda los desafíos únicos de la implementación en sistemas embebidos, asegurando un análisis espacial eficiente y decisiones precisas en tiempo real.

V. CONCLUSIONES

En conclusión, el artículo cumple de manera efectiva con su propósito de arrojar luz sobre el funcionamiento de los vehículos autónomos, destacando la complejidad inherente de los desafíos que estos enfrentan. Desde la necesidad crítica de reconocimiento espacial hasta la toma de decisiones en tiempo real, se evidencian los obstáculos significativos que surgen en el camino hacia la conducción autónoma segura y eficiente.

La presentación detallada de estrategias de co-diseño, especialmente aquellas que incorporan dispositivos FPGA y combinaciones eficientes de CPU/FPGA, se erige como un punto culminante. Estas estrategias no solo abordan los problemas fundamentales de velocidad de procesamiento, eficiencia y costos, sino que también demuestran ser catalizadores para soluciones mejoradas. La optimización de tareas críticas, la paralelización y la implementación específica en hardware se postulan como enfoques esenciales para superar los desafíos técnicos en la implementación de sistemas embebidos para vehículos autónomos.

La sección dedicada al Deep Learning y las Redes Neuronales, junto con la solución HALP de NVIDIA, subraya la importancia de tecnologías avanzadas para la comprensión y respuesta a patrones complejos en tiempo real. A pesar de la complejidad computacional, la implementación eficiente de estas tecnologías en sistemas embebidos se presenta como una realidad alcanzable mediante estrategias de co-diseño bien concebidas.

En resumen, el artículo ofrece una visión integral de cómo las estrategias de co-diseño pueden ser clave para el desarrollo de soluciones más efectivas en el ámbito de los vehículos autónomos. Al abordar los desafíos con enfoques innovadores y proporcionar soluciones tangibles, el artículo contribuye significativamente al avance y comprensión de la conducción autónoma, mostrando un camino hacia sistemas más eficientes, seguros y capaces.

REFERENCES

- [1] Shen Maying, Yin Hongxu, Molchanov Pavlo, Mao Lei, Liu Jianna and Alvarez Jose, "Structural Pruning via Latency-Saliency Knapsack". Advances in Neural Information Processing Systems, 2022.
- [2] D. Llopis Castelló, "Uso de vehículos autónomos y conectados," Departamento de Ingeniería e Infraestructura de los Transportes, Universitat Politècnica de València.
- [3] Quentin Cabanes. "New hardware platform-based deep learning co-design methodology for CPS prototyping : Objects recognition in autonomous vehicle case-study." Other [cs.OH]. Université Paris-Saclay, 2021.

- [4] J. L. Vázquez Gutiérrez, "ACELERADORES DE CÓDIGO BASADOS EN FPGA: QUÉ SON Y CÓMO HACERLOS (FPGA-based code accelerators: what are they and how to build one)," Trabajo de Fin de Grado, Universidad de Cantabria, Facultad de Ciencias, septiembre de 2016. Director: E. Villar Bonet, Co-Director: P. P. Sánchez Espeso.
- [5] C Collin, Anne, Siddiqi, Afreen, Imanishi, Yuto, Rebentisch, Eric, Tanimichi, Taisetsu et al. 2019. "Autonomous driving systems hardware and software architecture exploration: optimizing latency and cost under safety constraints." *Systems Engineering*, 23.