

Salto cuántico

22/04/2022 Cada vez más funciones de los vehículos están basadas en la inteligencia artificial. Sin embargo, los procesadores convencionales e incluso los gráficos están llegando a sus límites para realizar los cálculos necesarios para las redes neuronales. Este es un informe de Porsche Engineering sobre las nuevas tecnologías que acelerarán esos cálculos en el futuro.

La inteligencia artificial (IA) es una tecnología clave para la industria del automóvil. Por tanto, un hardware rápido es imprescindible para los complejos cálculos implicados en la gestión de la información. Al fin y al cabo, en el futuro solo con computadores de alto rendimiento será posible introducir nuevas funciones en la producción en serie. "La circulación autónoma es una de las aplicaciones de IA más exigentes", dijo el Dr. Joachim Schaper, Director Senior de IA y Big Data de Porsche Engineering. "Los algoritmos aprenden a partir de multitud de situaciones captadas en el tráfico real por los vehículos de prueba, mediante cámaras, radares u otros sensores".

Los centros de datos convencionales son cada vez más incapaces de hacer frente a la creciente demanda. "Ahora se necesitan días para entrenar una sola variante de una red neuronal", dijo Schaper. Así que, en su opinión, una cosa está clara: los fabricantes de automóviles necesitan nuevas tecnologías para los cálculos de IA que puedan ayudar a los algoritmos a aprender mucho más rápidamente. Para

lograrlo, es necesario ejecutar en paralelo el mayor número posible de multiplicaciones vectoriales en las complejas redes neuronales profundas (DNN), una tarea en la que están especializadas las unidades de procesamiento gráfico (GPU). Sin ellas, no habrían sido posibles los sorprendentes avances de la IA en los últimos años.

50 veces el tamaño de una GPU

Sin embargo, las tarjetas gráficas no se diseñaron originalmente para su aplicación en la IA, sino para procesar las imágenes de la forma más eficiente posible. Cada vez están más al límite cuando se trata de entrenar algoritmos para vehículos autónomos. Por tanto, se necesita hardware específico de IA para realizar esos cálculos. La empresa californiana Cerebras presentó una posible solución. Su Wafer Scale Engine (WSE) se adapta de forma óptima a los requisitos de las redes neuronales combinando toda la potencia de cálculo posible en un chip informático gigante. Es más de 50 veces mayor que un procesador gráfico normal y ofrece espacio para 850 000 núcleos de computación, más de 100 veces más que en una GPU actual de primer nivel.

Además, los ingenieros de Cerebras han conectado en red los núcleos de cálculo con líneas de datos de gran ancho de banda. Según el fabricante, la red del Wafer Scale Engine transporta 220 petabits por segundo. Cerebras también ha ampliado el cuello de botella dentro de las GPU: los datos viajan entre la memoria y la unidad de cálculo casi 10 000 veces más rápidamente que en las GPU de alto rendimiento, a 20 petabytes por segundo.

Para ahorrar aún más tiempo, Cerebras imita un truco del cerebro. En él, las neuronas solo funcionan cuando reciben señales de otras neuronas. Las numerosas conexiones que están inactivas en ese momento no consumen recursos. En cambio, en las DNN, la multiplicación de la matriz vectorial implica a menudo multiplicar por el número cero. Esto cuesta tiempo innecesariamente. Por ello, el motor Wafer Scale evita hacerlo. "Se filtran todos los ceros", escribe Cerebras en su libro blanco sobre el WSE. Así, el procesador solo ejecuta operaciones que producen un resultado distinto de cero.

Uno de los inconvenientes del chip es su elevada demanda de energía eléctrica, 23 kW, y que requiere refrigeración por agua. Por ello, Cerebras ha desarrollado su propia carcasa de servidor para su uso en centros de datos. El Wafer Scale Engine ya está siendo probando en los centros de datos de algunos institutos de investigación. El experto en IA Joachim Schaper cree que el gigantesco procesador californiano también podría acelerar el desarrollo del automóvil. "Teóricamente, con el Wafer Scale Engine una semana de aprendizaje podría reducirse a unas pocas horas", estima. "Sin embargo, la tecnología aún tiene que demostrarlo en pruebas empiricas".

Luz en lugar de electrones

Por muy inusual que sea este nuevo procesador también funciona con transistores convencionales, al igual que sus predecesores. En lugar de la electrónica, comparativamente lenta, empresas como

Lightelligence y Lightmatter quieren utilizar un medio mucho más rápido para los cálculos de IA: la luz. Para ello, en sus sedes de Boston están desarrollando procesadores ópticos. Así, las DNN podrían funcionar "al menos varios cientos de veces más rápido que las electrónicas", dicen los desarrolladores de Lightelligence.

Para ello, Lightelligence y Lightmatter utilizan el fenómeno de la interferencia. Cuando las ondas de luz se amplifican o anulan entre sí, forman un patrón de luz-oscuridad. Si se dirige la interferencia de una manera determinada, el nuevo patrón corresponde a la multiplicación vectorial del patrón anterior. Así que las ondas de luz pueden 'hacer matemáticas'. Para que esto sea funcional, los desarrolladores de Boston grabaron diminutas guías de luz en un chip de silicio. Como en un tejido, se cruzan varias veces. En los cruces se producen interferencias. Entre medias, unos diminutos elementos calefactores regulan el índice de refracción de la guía luminosa, permitiendo que las ondas de luz se desplacen unas contra otras. Esto permite controlar su interferencia y realizar multiplicaciones vectoriales matriciales.

Sin embargo, las empresas de Boston no prescinden totalmente de la electrónica. Combinan sus ordenadores de luz con componentes electrónicos convencionales que almacenan datos y realizan todos los cálculos, excepto las multiplicaciones vectoriales matriciales. Entre ellos se encuentran, por ejemplo, las funciones de activación no lineal que modifican los valores de salida de cada neurona antes de pasar a la siguiente capa.

Al combinar la computación óptica y la digital, las DNN pueden calcular con extrema rapidez. "Su principal ventaja es la baja latencia", dijo Lindsey Hunt, portavoz de Lightelligence. Esto permite, por ejemplo, que la DNN detecte más rápidamente los sujetos en las imágenes, como peatones o personas que circulan en sillas de ruedas eléctricas. En la circulación autónoma, esto podría significar reacciones más rápidas ante situaciones críticas. "Además, el sistema óptico toma más decisiones por vatio de energía eléctrica", afirma Hunt. Esto es especialmente importante, ya que el aumento de la potencia de cálculo en los vehículos se produce cada vez más a expensas del ahorro de combustible y la autonomía.

Las soluciones de Lightmatter y Lightelligence pueden insertarse como módulos en los ordenadores convencionales para acelerar los cálculos de IA, como las tarjetas gráficas. En principio, también podrían integrarse en vehículos, por ejemplo, para implementar funciones autónomas. "Nuestra tecnología es muy adecuada para servir de motor de inferencia en un auto autónomo", dijo Lindsey Hunt. El experto en IA Schaper tiene una opinión similar: "Si Lightelligence consigue construir componentes adecuados para los automóviles, esto podría acelerar enormemente la introducción de funciones complejas de IA en los vehículos". La tecnología ya está lista para el mercado: la empresa está planificando sus primeras pruebas piloto con clientes en este año.

El computador cuántico como impulsor de la IA

Los computadores cuánticos están algo más lejos de la aplicación práctica. También acelerarán los cálculos de la IA porque pueden procesar grandes cantidades de datos en paralelo. Para ello, trabajan con los llamados 'qubits'. A diferencia de la unidad clásica de información, el bit, un qubit puede

representar los dos valores binarios 0 y 1 simultáneamente. Los dos números coexisten en un estado de superposición que solo es posible en la mecánica cuántica.

Los computadores cuánticos podrían dar un impulso a la inteligencia artificial a la hora de clasificar objetos, por ejemplo en el tráfico. Hay muchas clases diferentes de objetos; entre ellos, bicicletas, autos, peatones, señales de tránsito, carreteras mojadas y secas. Se diferencian en muchas propiedades, por lo que los expertos hablan de 'reconocimiento de patrones en espacios de dimensión superior'.

"Cuanto más complicados son los patrones, más difícil es para los computadores convencionales distinguir las clases", dijo Heike Riel, que dirige la investigación cuántica de IBM en Europa y África. Esto se debe a que, con cada dimensión, resulta más costoso calcular la similitud de dos objetos: ¿Qué tan parecidos son una persona en una silla de ruedas eléctrica o con un andador al cruzar una calle? Los computadores cuánticos pueden trabajar eficazmente en espacios de dimensión superior en comparación con los ordenadores convencionales. Para ciertos problemas, esta propiedad podría ser útil y dar lugar a que algunos problemas sean resueltos más rápidamente con la ayuda de los computadores cuánticos que con los computadores convencionales de alto rendimiento.

Los investigadores de IBM han analizado los modelos estadísticos que pueden ser entrenados para clasificar datos. Los primeros resultados sugieren que, para determinados conjuntos de datos, los modelos cuánticos elegidos de forma inteligente funcionan mejor que los métodos convencionales. Los modelos cuánticos son más fáciles de entrenar y parecen tener mayor capacidad, lo que les permite aprender relaciones más complicadas.

Riel admite que, aunque los computadores cuánticos actuales pueden utilizarse para probar estos algoritmos, todavía no tienen ventaja sobre los computadores convencionales. Sin embargo, el desarrollo de los computadores cuánticos avanza rápidamente. Tanto el número de qubits como su calidad no dejan de aumentar. Otro factor importante es la velocidad, medida en operaciones de capa de circuito por segundo (CLOPS). Este número indica cuántos circuitos cuánticos pueden ejecutarse en el computador cuántico por unidad de tiempo. Es uno de los tres criterios de rendimiento importantes de un computador cuántico: escalabilidad, calidad y velocidad.

En un futuro previsible, debería ser posible demostrar la superioridad de los computadores cuánticos para determinadas aplicaciones; es decir, resuelven los problemas con mayor rapidez, eficacia y precisión que un computador convencional. Pero la construcción de un computador cuántico potente, con corrección de errores y de uso general todavía llevará algún tiempo. Los expertos calculan que tardará al menos otros 10 años. Aun así, la espera podría merecer la pena. Al igual que los procesadores ópticos o las nuevas arquitecturas de los ordenadores electrónicos, los computadores cuánticos podrían ser la clave de la movilidad del futuro.

En resumen

Cuando se trata de cálculos de IA, no solo los microprocesadores convencionales, sino también los

gráficos están llegando a sus límites. Por ello, empresas e investigadores de todo el mundo trabajan en nuevas soluciones. Los chips en formato de oblea (wafer) y los computadores ópticos están a punto de hacerse realidad. Dentro de unos años, éstos podrían complementarse con computadores cuánticos para cálculos especialmente exigentes.

Información

Artículo publicado en la edición número 1/2022 de la revista Porsche Engineering.

Texto: Christian Meier

Derechos de autor: las imágenes y el sonido aquí publicados tienen derechos de autor de Dr. Ing. h.c. F. Porsche AG, Alemania, u otras personas. Está prohibida la reproducir total o parcial sin autorización escrita de Dr. Ing. h.c. F. Porsche AG. Por favor, contacte con newsroom@porsche.com para más información.

Link Collection

Link to this article
https://newsroom.porsche.com/es/2022/tecnologia/PLA-es-porsche-engineering-microprocesador-chip-oblea-ordenador-luz-optico-cuantico-28074.html

External Links
https://www.porscheengineering.com/peg/en/

Dr. Ing. h.c. F. Porsche Aktiengesellschaft
Porscheplatz 1
70435 Stuttgart