

GPU vs CPU, diferencias y similitudes

GPU vs CPU, diferencias y similitudes. CPUs y GPUs tienen mucho en común. Ambos son microprocesadores basados en silicio. Al mismo tiempo, son sustancialmente diferentes y se implementan para diferentes roles.

¿Qué son las CPU y las GPU?

Una CPU (unidad central de procesamiento) a menudo se llama el «cerebro» o el «corazón» de una computadora. Se requiere para ejecutar la mayor parte del software de ingeniería y de oficina. Sin embargo, hay una multitud de tareas que pueden abrumar el procesador central de una computadora. Eso es cuando usar GPU se vuelve esencial para la computación.

Una GPU (unidad de procesamiento de gráficos) es un tipo especializado de microprocesador, diseñado principalmente para la representación rápida de imágenes. Las GPU aparecieron como una respuesta a aplicaciones gráficamente intensas que suponen una carga para la CPU y un rendimiento del equipo degradado. Se convirtieron en una forma de descargar esas tareas de las CPU, pero los procesadores gráficos modernos son lo suficientemente potentes como para realizar cálculos matemáticos rápidos para muchos otros propósitos además de la representación.

¿Cuál es la diferencia?

CPUs y GPUs procesan tareas de diferentes maneras. En cuanto a las interrelaciones, a menudo se comparan con el cerebro y la fuerza muscular. Una CPU (el cerebro) puede funcionar en una variedad de cálculos diferentes, mientras que una GPU (la fuerza física) es mejor para enfocar todas las habilidades de

computación en una tarea específica. Esto se debe a que una CPU consta de unos pocos núcleos (hasta 24) optimizados para el procesamiento secuencial en serie. Está diseñado para maximizar el rendimiento de una sola tarea dentro de un trabajo; Sin embargo, la gama de tareas es amplia.

Por otro lado, una GPU utiliza miles de núcleos más pequeños y más eficientes para una arquitectura paralela masiva dirigida a manejar múltiples funciones al mismo tiempo.

Las GPU modernas proporcionan una potencia de procesamiento, ancho de banda de memoria y eficiencia superiores a las de su CPU. Son 50–100 veces más rápidos en tareas que requieren múltiples procesos paralelos, como el aprendizaje automático y el análisis de big data.

¿Qué problemas son adecuados para abordar GPUs?

La computación de GPU se define como el uso de una GPU junto con una CPU para acelerar las aplicaciones científicas, analíticas, de ingeniería, de consumo y empresariales.

Durante muchos años, las GPU han impulsado la visualización de imágenes y movimientos en pantallas de computadoras, pero son técnicamente capaces de hacer más. Los procesadores gráficos se ponen en juego cuando se necesitan cálculos masivos en una sola tarea.

Esa tarea puede incluir:

Juegos

Una unidad de procesamiento de gráficos es esencial para la representación rápida y gráfica del mundo de los juegos. La renderización de efectos especiales y sofisticados gráficos 3D en tiempo real requiere un gran poder de cómputo. Las tareas de los juegos modernos se vuelven demasiado pesadas para la

solución de gráficos de CPU. Los juegos incluso dieron un paso más allá con la realidad virtual, que es tan creíble porque las GPU pueden procesar y mantener imágenes realistas rápidamente con la iluminación y el sombreado adecuados.



Visualización 3D

Las GPU impulsan el rendimiento de la ventana gráfica en aplicaciones de visualización 3D, como el diseño asistido por computadora (CAD). El software que le permite visualizar objetos en 3 dimensiones se basa en las GPU para dibujar esos modelos en tiempo real a medida que los gira o los mueve.

Procesamiento de imágenes

Las GPU pueden procesar con precisión millones de imágenes para encontrar diferencias y similitudes. Esta capacidad se usa ampliamente en industrias como el control de fronteras, la seguridad y el procesamiento médico de rayos X. Por ejemplo, en 2010, el ejército de EE. UU. Conectó más de 1.700 sistemas Sony PlayStation 3™ para procesar imágenes satelitales de alta resolución con mayor rapidez.

Big Data

Con miles de núcleos computacionales y un rendimiento de aplicación de 10–100x en comparación con las CPU solo, las unidades gráficas son la opción para procesar Big data para científicos e industrias. Las GPU se utilizan para representar datos como visualización interactiva y se integran con otros conjuntos de datos para explorar el volumen y la velocidad de los datos. Por ejemplo, ahora podemos potenciar el mapeo de genes procesando datos y analizando las co-varianzas para entender la relación entre diferentes combinaciones de genes.



Aprendizaje profundo de la máquina

El aprendizaje automático ha existido desde hace algún tiempo, pero la informática de GPU potente y eficiente lo ha elevado a un nuevo nivel. El aprendizaje profundo es el uso de redes neuronales sofisticadas para crear sistemas que puedan realizar la detección de características a partir de cantidades masivas de datos de entrenamiento sin etiquetar. Las GPU pueden procesar toneladas de datos de entrenamiento y entrenar redes neuronales en áreas como análisis de imágenes y

video, reconocimiento de voz y procesamiento de lenguaje natural, autos de auto manejo, visión por computadora y mucho más.

Las GPU no son reemplazos para la arquitectura de la CPU. Más bien, son poderosos aceleradores para la infraestructura existente. La computación acelerada por GPU descarga partes de la aplicación con uso intensivo de cómputo a la GPU, mientras que el resto del código aún se ejecuta en la CPU. Desde la perspectiva de un usuario, las aplicaciones se ejecutan mucho más rápido. Si bien la computación de propósito general sigue siendo el dominio de la CPU, las GPU son la columna vertebral del hardware de casi todas las aplicaciones computacionales intensivas.

¿Por qué GPUs?

No es ningún secreto en la comunidad de Deep Learning que las máquinas y los clusters habilitados para GPU acelerarán dramáticamente el tiempo que toma entrenar redes neuronales. Sin embargo, una de las mejores características de los algoritmos de redes neuronales es su extensibilidad a la paralelización. Podemos dividir las n instancias y m actualizaciones de dimensión en el tamaño del lote en c núcleos diferentes en nuestra GPU.

Las GPU NVIDIA suelen tener más de 1.000 núcleos, en comparación con las CPU que son increíblemente rápidas, pero solo tienen entre 4 y 8 núcleos. ¿Qué significa esto para nuestro entrenamiento de redes neuronales?

Estos núcleos nos permiten dividir nuestro lote (n / c) y ejecutar el proceso de capacitación en cada uno de estos núcleos en paralelo. Esto es posible debido a la forma en que Gradient Descent actualiza los parámetros de peso en la red. En lugar de actualizar instantáneamente los pesos después de cada paso de entrenamiento, GD agrega el error para el lote y luego actualiza los pesos después de que se haya calculado la

actualización para todas las instancias en el lote. Además, la actualización de los pesos se puede dividir, ya que estas operaciones también se ejecutan de forma independiente.

Leer también: [La computación acelerada por GPU se abre paso en el centro de datos](#); [Por qué es tan importante el análisis Big Data](#) ; [Edge computing qué es, para qué sirve](#)