



繪圖處理器

資工二 111210507 林育民

GPU 圖形處理器

圖形處理單元 (GPU) 是一種可高速執行數學計算的電子電路。圖形轉譯、機器學習 和視訊編輯等運算任務，需要在大型資料集上運用類似的數學運算。GPU 的設計可讓它平行地對多個資料值執行相同的操作。這可提高其對許多運算密集型任務的處理效率。



GPU 技術的演變

- 1999 年，Nvidia 率先推出單晶片 GeForce 256 GPU。21 世紀 00 年代和 10 年代標誌著一個成長時代，期間 GPU 融入了光線追蹤、網格陰影和硬體曲面細分等功能。這推進了影像產生的處理，同時提升了圖形效能。
- 2007 年，Nvidia 發佈了 CUDA，這是在 GPU 上提供平行處理的軟體層。大約在此期間，顯而易見的是，GPU 在執行高度特定的任務方面非常高效。它們擅長完成需要大量處理能力才能實現特定結果的任務。
- 當 Nvidia 發佈 CUDA 時，它向更廣泛的受眾開放 GPU 程式設計。然後，開發人員可以為各種不同的運算密集型實際應用程式設計 GPU 技術。GPU 運算開始變得更加主流化。

GPU 和 CPU區別

- CPU 處理完整的系統控制以及管理和一般用途任務。相反的，GPU 處理運算密集型任務，例如視訊編輯或機器學習
- GPU有很多核心，能在同時處理大量小型任務，適合平行任務
CPU有少量的高活躍核，適合處理小量但複雜的任務。
- GPU在處理平行計算時，優勢明顯，速度遠高於CPU。CPU在計算單一任務時，速度是優於GPU的。

GPU vs CPU比較圖

GPU vs CPU		
特性	GPU（圖形處理器）	CPU（中央處理器）
架構	大量簡單的核心，專為並行運算	少量強大的核心，適合串行運算
計算能力	浮點運算能力強	整數運算能力較佳
記憶體存取	有小容量的高速緩存（如寄存器和共享記憶體），存取速度快； 全局記憶體容量大，但存取速度較慢，頻寬高。	有大容量的高速緩存（L1、L2、L3），存取速度非常快； 主記憶體容量大，存取速度相對於快取記憶體慢。
功耗	一般功耗較高	相對功耗較低
應用領域	圖形渲染、深度學習、科學計算	一般運算、系統管理、IO等
平行計算	大量線程，核心適合高度並行	核心數量相對較少，並行能力差

CUDA繪圖處理器

- CUDA 是一種由 NVIDIA 公司開發的運算平台和程式設計模型，全稱為 Compute Unified Device Architecture，旨在充分利用 GPU 的高度平行運算能力，加速各種複雜的運算任務。CUDA 程式設計模型採用異質平行計算架構，將工作切分成數以百萬計的更小的線程，並行執行於 GPU 上的數以千計的 CUDA 核心上。



CUDA工作原理

- 每一個**CUDA**核心都可以獨立運行一段稱為核函數的程式代碼，這些核函數會在**GPU**內部的多個處理單元上並行執行
- **CUDA**採用任務分拆的方式實現大規模並行運算。它將原本的計算任務拆分為數百甚至數千個小型子任務，再指派給**GPU**中數以千計的**CUDA**核心同時執行。

CUDA的應用領域

- 數據科學家可利用**CUDA**強勁的**GPU**加速能力，高效完成數據清理、特徵工程和模型訓練等關鍵分析流程，大幅縮短了整體分析時間。這一優勢對需要處理大規模複雜數據集的金融、行銷、科研等領域造就龐大的商機

結語

- 在AI領域，**CUDA**和**GPU**的高效能對於大規模神經網絡訓練至關重要，促進了自然語言處理、電腦視覺和自動化系統等領域的發展。
- 在未來發展的方面:隨著**GPU**硬件性能的持續提升及**CUDA**等軟件平台的優化，未來的圖形處理技術將在更多領域內展現其優勢，例如量子計算模擬、生物醫學分析和自動駕駛系統。這些應用將推動人類科技邁向更高的層次。