SAMSUNG

三星電子為佈局 AI 深度學習推出高速、低功耗 NPU 解決方案

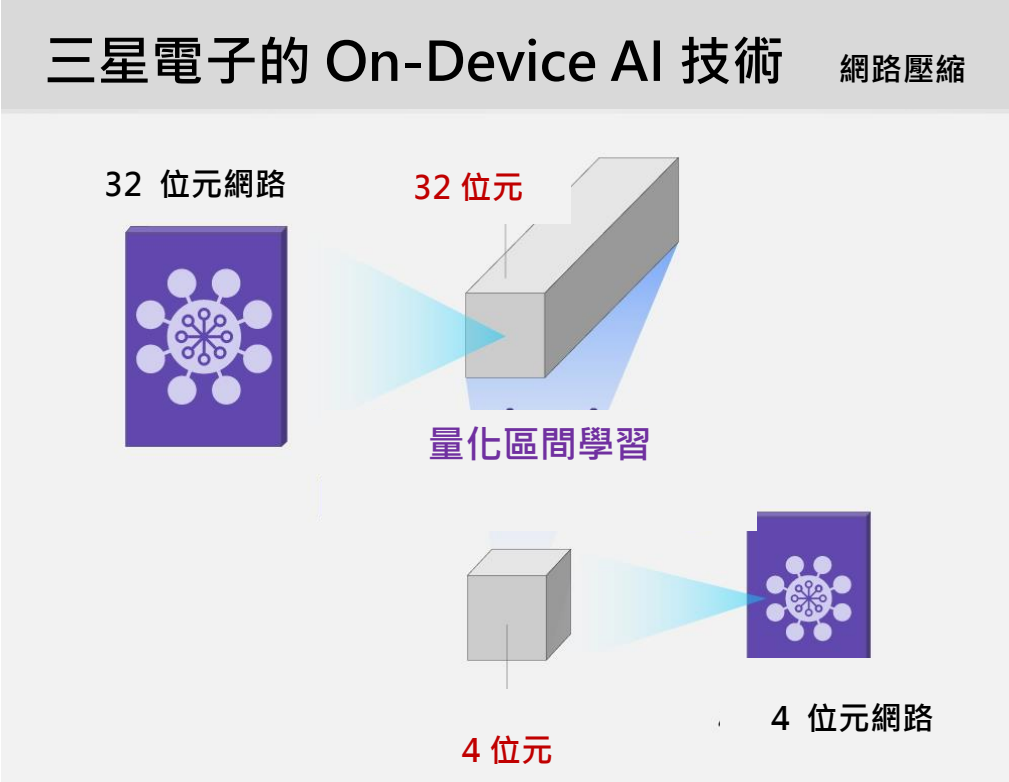
深度學習演算法是人工智慧(AI)的核心要素，藉由這演算過程，電腦能像人類思考和學習。神經處理單元(NPU)是一種針對深度學習演算的優化處理器，讓數以千計的運算獲得高效率的同步處理。

三星電子上月宣佈，將在 2030 年以前擴大 NPU 專有技術的研發，藉以強化其在全球系統半導體產業的領導地位。三星電子近日在全球頂尖電腦視覺領域的學術會議 -「電腦視覺和模式識別(CVPR)」會議上，暢談未來發展的新願景。

三星電子在 CVPR 上引述一篇論文《Learning to Quantize Deep Networks by Optimizing Quantization Intervals With Task Loss》，闡述其對 On-Device AI 輕量級演算法的研發投入。On-Device AI 技術可以直接在裝置端運算和處理資料。三星電子最新的演算方案大幅進化，比現有的演算法輕量 4 倍、速度快 8 倍以上，展現低功耗、高速運算的絕佳優勢，被視為解決潛在問題的利器。

簡化深度學習過程

三星先進技術研究院(SAIT)日前宣佈，其已成功開發 On-Device AI 輕量級技術，其運算速度比現有 32 位元伺服器深度學習數據快 8 倍。透過將數據重新分組為 4 位元，並維持資料準確的識別性，這個全新的深度學習演算法，比現有方案更快、更節能地實現同步處理。



對於影響深度學習整體成效的重要資料，三星電子的最新 On-Device AI 處理技術，透過「學習」來決定區間的大小。這項「Quantization[註一] Interval Learning (QIL)」技術藉由重新分組，使數據小於原始的位元數，以維持數據的準確性。SAIT 的實驗結果證明，原始 32 位元區間的伺服器深度學習演算法，在量化至 4 位元以下的區間之後，因此比現有的其它解決方案具有更高的精確度。

當深度學習運算的資料被分組為 4 位元以下時，除了加法和乘法的算術計算外，還能進行「和」及「或」的邏輯運算。這表示使用 QIL 處理程序的運算，可以獲得與現有程序相同的結果，卻只需要 1/40 至 1/120 甚至更少的電晶體[註二]。

由於該系統需要較少的硬體和電力，因此可以直接安裝在影像資料的裝置中，或是指紋感應器中，更勝於將處理後的數據傳輸至其他必要的端點。

SAMSUNG

AI 處理和深度學習的未來

On-Device AI 處理技術將有助於提升三星電子的半導體實力，也有利於強化其在 AI 時代的核心競爭力。不同於使用 AI 雲端伺服器的 AI 服務，On-Device AI 能在裝置端直接運算其獲取的資料。

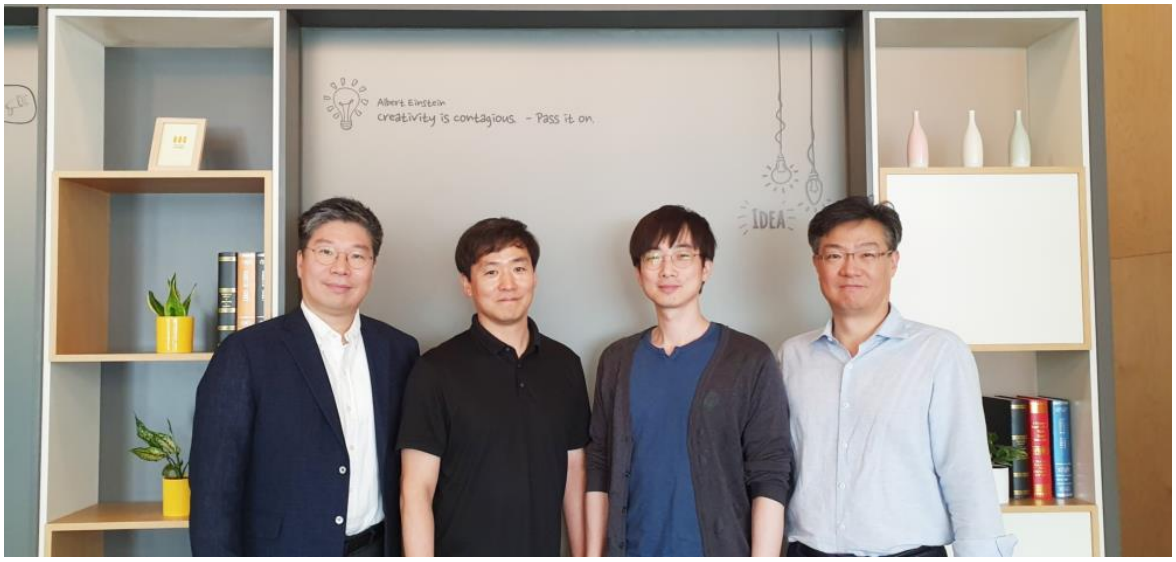
On-Device AI 技術能降低雲端伺服器的建置成本，因為它本身具備運算能力，同時能為虛擬實境、自動駕駛等應用情境，帶來既快速又穩定的性能表現。此外，On-Device AI 技術能將裝置身份認證用的個人生物資訊，例如指紋、虹膜和臉部掃描等，安全地儲存在行動裝置上。

三星電子副總裁暨 SAIT 電腦視覺實驗室負責人 Chang-Kyu Choi 談到：「在未來的世界裡，AI 將主宰人們生活中的所有裝置和感應器。三星電子的 On-Device AI 技術，是深度學習且低功耗

與高速解決方案，為未來世界鋪路。三星也將擴大應用於記憶體、處理器和感應器，以及其它次世代系統半導體市場。」

On-Device AI 技術的核心功能，能發揮高速運算且十分省電。三星電子的第一個解決方案是去年推出的 Exynos 9(9820)，這個系統晶片(SoC)搭載專有技術 Samsung NPU，讓行動裝置得以執行 AI 運算，不需仰賴任何的外部雲端伺服器。

各大企業紛紛將注意力轉到 On-Device AI 技術上。三星電子計畫在不久的將來，將該演算法的應用從行動 SoC 延伸至記憶體、感應器等解決方案，藉以強化其在 AI 技術的領先地位。



三星電子 On-Device AI 輕量化演算法技術開發的四大主將。

(從左到右)：來自於三星電子先進技術研究院(SAIT)的 Jae-Joon Han、Chang-Young Son、Sang-Il Jung、Chang-Kyu Choi

註一：Quantization 量化是減少數據位元數的一種過程，它將既有數據分割成若干個有限區段，分割後的數據能以某個位元數表示，且各區段中的數據具有相同的值。

註二：電晶體是一種藉由放大或開關作用，控制半導體電流或電壓的裝置。