

Arm Ethos-U65: 在人工智慧裝置的新世界中帶動創新



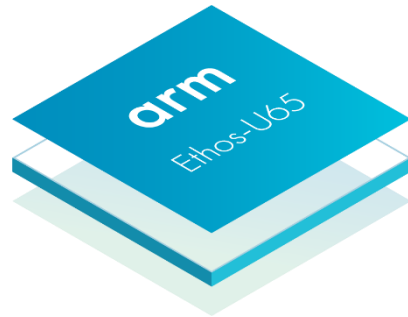
作者：Tanuj Arora

邊緣與終端人工智慧(AI)今日正在全球爆發性發展，對我們的日常生活帶來意義深遠的影響。AI 技術正運用在公共安全、提升零售體驗、交通運輸，以及其它龐大且快速成長的各種場景。從智慧門鎖到家庭電子產品等消費應用，AI 正在進一場革命性的變革。在基礎設施方面，AI 正使用在如資料層(data plane)優化與電源管理等越來越多的領域。AI 真的是無所不在。

在今年年初，我們發布了 [Arm 機器學習\(ML\)產品組合](#) 的新力軍，以促成終端極度低功耗的機器學習推論。與上一世代應用 Cortex-M CPU 相比，[Arm Cortex-M55 CPU](#) 以及全球第一個微型神經網路處理器(NPU) [Arm Ethos-U55 NPU](#) 的組合，ML 效能大大提升了 480 倍。這個組合可以讓裝置在終端裝置上運行神經網路推論，毋需把大量的數據傳到雲端。把數據留在裝置上處理，不但可以讓系統的回應性更高，而且更加可靠、更安全，並且保有更多的隱私。Ethos-U55 的成功讓我們相信人類正朝向一個前所未見在裝置上的 AI 發展的未來快速邁進中。

利用全新的 Arm Ethos-U65 將 AI 擴展到新的裝置

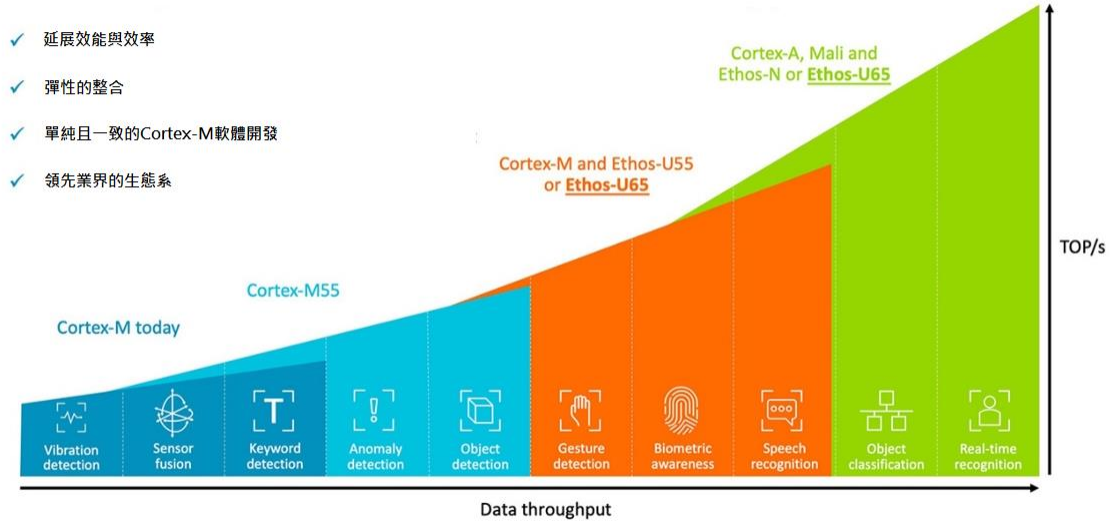
為了促成更多創新、並把 AI 應用擴展到更多的裝置中，Arm 宣布推出最新版本的 Ethos 產品線，也就是為高效能嵌入式裝置與次系統提供神經網路加速的 [Arm Ethos-U65](#) 微型 NPU。Ethos-U65 保有 Ethos-U55 的功耗效率，同時把應用性延伸至 Arm Cortex-A 與 Arm Neoverse 架構系統。儘管 Ethos-U65 可以與 Cortex-A 或 Neoverse CPU 一起使用，但它格外適合搭配具先進向量能力的 Arm CPU，例如 [Arm Cortex-A55](#)。



微型尺寸、超高效率

Ethos-U55 導入我們首個微型 NPU 架構，它可以讓神經網路加速在面積極小的地方，以低電力消耗進行。Ethos-U65 以 Ethos-U55 的成功為基礎，同時維持對功耗效率的專注，把應用性延伸到 Arm Cortex-A 與 Neoverse 架構系統。Ethos-U65 為需求更高的應用提供全新更高的效能點，達到 1 TOPs，可以賦予各種裝置新能力，如高解析智慧相機、智慧家庭解決方案，以及甚至如頻寬與電力管理次系統等基礎設施應用。

在AI應用的新世界中帶動創新



延展的網路支援

Ethos-U 產品線設計的基礎是它對 ML 運算子的原生支援，這使它能夠完全在 NPU 上執行最流行的神經網路，而不會發生運算子後降到 CPU 的情形。以 Ethos-U65 來說，這個運算子支援已經進一步完成更新與擴展。不過，在後降 CPU 仍是必須的情況下，這些運算子通常還是會透過如 Arm Compute Library 計算庫與 CMSIS-NN 等 Arm 高度優化軟體完成加速。

Ethos-U 的網路支援

- With the supported operator set, Ethos-U can completely execute most of the popular networks today. For ex:



...and many more

- For networks that cannot be executed on Ethos-U completely, the operators unsupported by Ethos-U fallback to the attached Cortex-M CPU
 - These are accelerated through CMSIS-NN library

Ethos-U65 提供 256 與 512 MACs/週期等兩種不同的組態配置。它包括一個可以為權重受限網路帶來更佳頻寬的雙 AXI(高級可擴展介面)。

- 對 SRAM (128 位元 AXI)與 DRAM (128 位元 AXI)提供支援。對 A 類記憶體系統提供支援

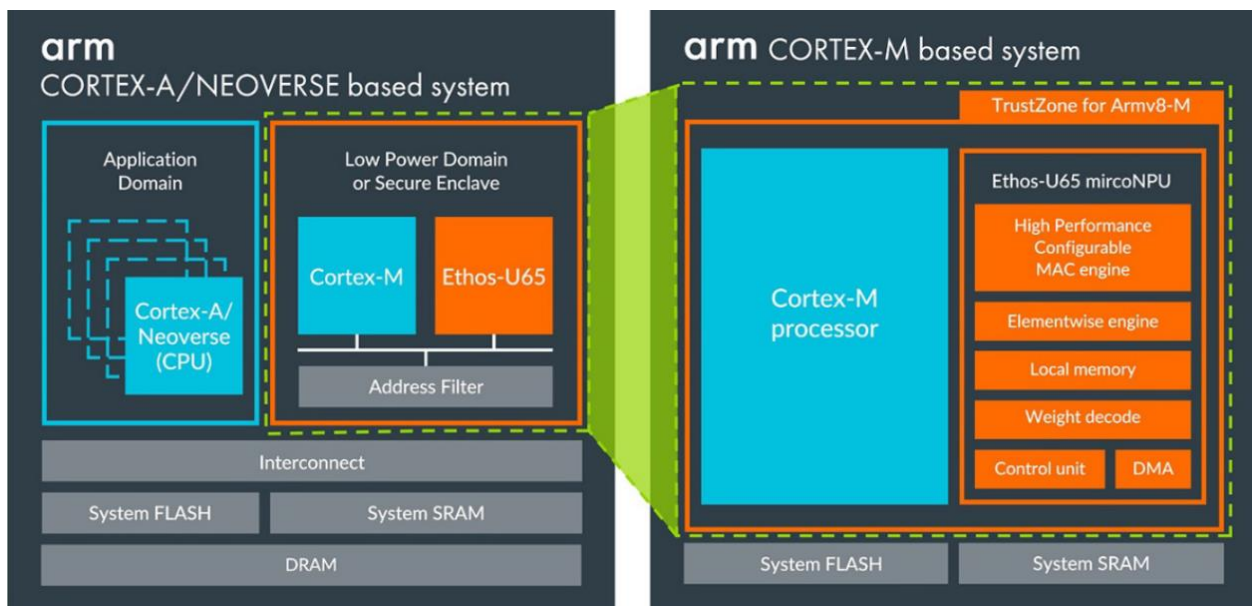
對於這些較為複雜的系統，與 Ethos-U55 相比，Ethos-U65 平均可以帶來 150%的網路效能（推論/秒）提升。

針對您的 ML 需求選擇正確的系統

Ethos-U65 設計就是用來與 DRAM 架構的系統一起使用，可以帶來更高可用的頻寬。這讓 Ethos-U65 可以用在更多類型的嵌入式系統中：

- 更高效能的 Cortex-A 與 Neoverse 架構的系統單晶片，或是
- 供電池驅動裝置（配備或沒有配備 DRAM)使用的低功耗 Cortex-M 架構的系統單晶片

適用於 Ethos-U65 的不同應用



在 Cortex-A 或 Neoverse 架構的系統中使用 Ethos-U65 時，軟體流與 Ethos-U55 一樣，都是使用 TFLmicro runtime。TFLmicro 堆疊利用伴隨 NPU 的 Cortex-M 處理器運行，同時處理所有微型 NPU 本身無法執行的卸載運算子。

Ethos-U65 一個顯著的優點，是可以促成任何大小網路的執行。有了 Ethos-U65，您可以高效率地加速更大型的網路，促成包括即時物件檢測、分類與辨識等應用。

此外，有了 Ethos-U65，所有利用 Arm 處理器打造 ML 應用進行投資，不但不會流失，並且可以持續再使用，原因是 Ethos-U65 與 Ethos-U55 使用同樣的軟體與工具。除了 [統一的軟體與工具](#)，Arm Ethos-U 微型 NPU 還獲得豐富與生氣蓬勃的合作夥伴生態系支援，提供橫跨音訊與視覺使用場景的各種解決方案，包括語音辨識、影像分類與物件檢測。

加速 AI 能力進入邊緣與終端裝置

在邊緣與終端裝置內提供更多的 AI 能力，可以打開難以想像的創新、創造與效率大門，並引導我們開發驚人的未來產品。我最近閱讀過的一篇文章宣稱：「AI 將與電力一樣讓一切改觀。」對我個人而言，AI 是一個世代只出現一次的運算改變，會大大改觀一切，從雲端伺服器、智慧家庭解決方案，到最小的物聯網裝置。機會相當的龐大，市場也對所有人敞開，就讓我們拭目以待 Ethos-U65 NPU 可以促成哪些新應用。

- 點擊此(<https://bit.ly/2TixMO9>) 以瞭解更多有關 [Ethos-U65](#) 的資訊。
- 點擊 Arm 相關新聞 (<https://bit.ly/3m66cVx>) 以閱覽更多新聞發布，以及我們與恩智浦半導體(NXP)有關 Ethos-U65 的技術合作。



All brand names or product names are the property of their respective holders. Neither the whole nor any part of the information contained in, or the product described in, this document may be adapted or reproduced in any material form except with the prior written permission of the copyright holder. The product described in this document is subject to continuous developments and improvements. All particulars of the product and its use contained in this document are given in good faith. All warranties implied or expressed, including but not limited to implied warranties of satisfactory quality or fitness for purpose are excluded. This document is intended only to provide information to the reader about the product. To the extent permitted by local laws Arm shall not be liable for any loss or damage arising from the use of any information in this document or any error or omission in such information.