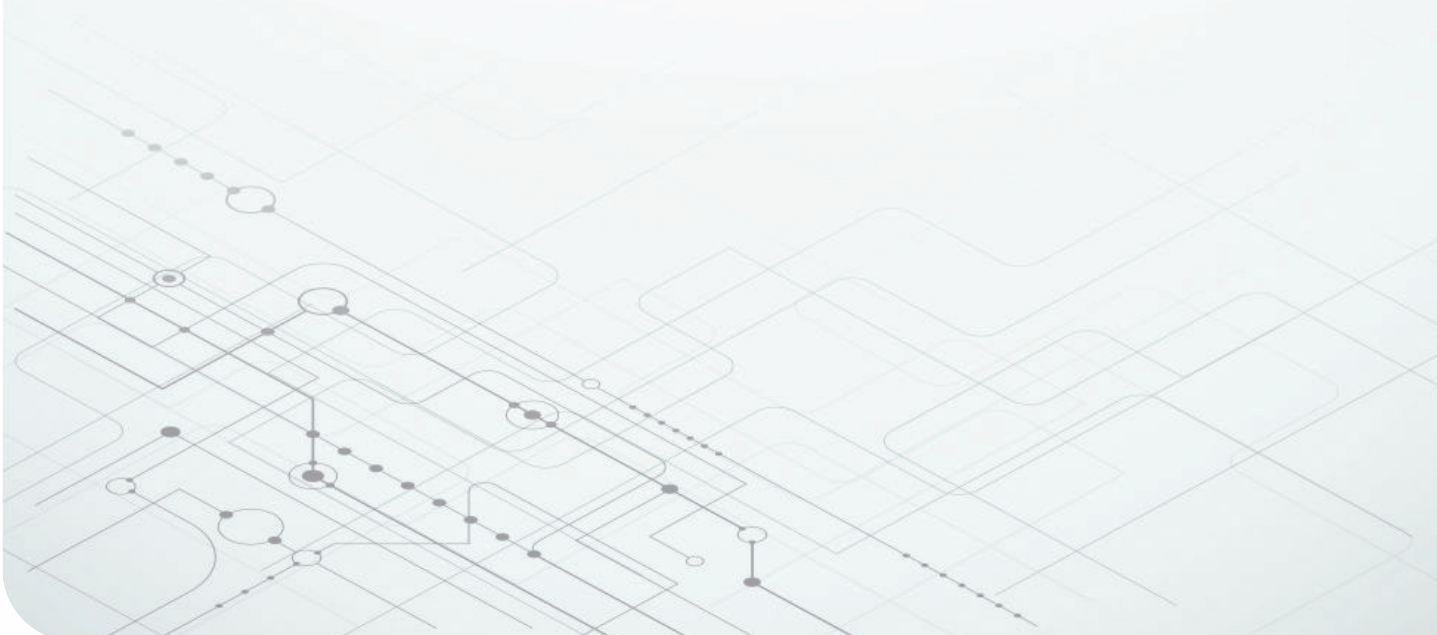


以高度整合處理器設計高效邊緣 AI 系統



Manisha Agrawal
Product Marketing
Jacinto™ Processors



自動化從工廠不斷擴展到大門。

摘要

摘要本白皮書將說明打造高效邊緣人工智慧 (AI) 系統的要求，以及視覺 AI 處理器如何藉由異質架構和可擴充 AI 性能，幫助實現最佳性能。

1 定義邊緣 AI
定義頂尖邊緣人工智慧。許多類型的系統都可從邊緣 AI 處理中獲益。

2 什麼是高效邊緣 AI 系統？
什麼是實際的邊緣 AI 系統？考慮哪些架構和核心最適合完成系統所需工作。

3 以 TI 視覺處理器設計邊緣 AI 系統
設計搭載視覺 AI 處理器的邊緣 AI 系統，如 TDA4 和 AM6xA 晶片系統 (SoC)。這些 SoC 可在低功耗與低系統 BOM 成本下，提供可擴充的輸送量和運算性能。

簡介

當消費者在線上訂購產品時，自動化可為程序中的每個步驟提高效率，不管是製造原料、提高倉庫生產力還是促進宅配，有時只需要幾小時即可送達客戶家中。若想在自動化方面持續進步，將需要更好的機器感知和智慧 and 更少錯誤，此時只要將人工智慧 (AI) 引進邊緣裝置即可達成此目標。

若要打造更快速、更聰明且更準確的系統，需要從更多感測器獲得更多資料，並需增加處理能力。但資料及運算增加也會對系統性能及其功耗和成本要求帶來挑戰。若想實現系統最佳化與縮短開發週期時間，必須以較實際的方式來設計邊緣 AI 系統。

定義邊緣 AI

邊緣 AI 以本機裝置處理 AI 演算法取代雲端處理，改變以深度神經網路 (DNN) 為主要演算法元件的工業和汽車應用。邊緣 AI 應用需要高速與低功耗處理，以及應用及其工作特有的進階整合，才能在尺寸受限、功率和散熱受限且

成本受限的環境中有效運作。圖 1 說明運用邊緣 AI 處理來提升性能和效率的部分應用。例如使用視覺輸入的邊緣 AI 系統可在生產線上以單一攝影機進行品質控制，或以多部攝影機支援汽車或行動機器人的功能安全。



圖 1. 邊緣智慧存在許多各種應用中

邊緣 AI 系統可協助改善倉庫與工廠效率，讓城市、建築與農業更安全、更有效率，並讓居家與零售環境更加智慧化。讓我們來看看幾個需要高效邊緣 AI 處理的系統：

- 1. 先進駕駛輔助系統 (ADAS)。** ADAS 技術可讓您深入瞭解車輛周圍環境，讓駕駛體驗更方便、更輕鬆且更安全。多數 ADAS 功能都是視覺型系統，從多部攝影機感測器取得高解析度輸入，並使用深度學習與電腦視覺演算法來解讀 ADAS 技術，以提供車輛周圍環境的資訊洞見，讓駕駛體驗更方便、更輕鬆且更安全。多數 ADAS 功能都是視覺型系統，從多部攝影機感測器取得高解析度輸入，並使用深度學習與電腦視覺演算法來解讀 ADAS 技術。
- 2. 自主行動機器人和無人機。** 就市售可用機器人來說，晶片系統 (SoC) 必須以高速與低功耗處理複雜的感知與導航堆疊，並需具備最佳系統成本。SoC 也必須為影像消除彎曲、立體聲深度評估、調整、影像金字塔生成與深度學習等運算密集工作進行卸載，以實現最大系統效率。
- 3. 智慧型購物車。** 智慧型購物車可計算購物車內物品的訂單總計、建議購物清單品項，並讓消費者能對購物車中的雜貨進行結帳，讓客戶享有更客製化的購物體驗，並可跳過結帳時的排隊人龍。大多數智慧型購物車都有多個視覺感測器以攝影機自動偵測物品，電腦

視覺智慧型購物車則可計算放入購物車品項的訂單總計、建議購物清單項目，並讓消費者能對購物車中的雜貨進行結帳，讓客戶享有更客製化的購物體驗，並可跳過結帳時的排隊人龍。

4. **邊緣 AI 運算盒**。邊緣 AI 運算盒是運用在零售自動化、工廠監控和建築監控系統的攝影機系統智慧延伸。儘管有尺寸限制、功率和散熱方面的挑戰，高傳輸速率 AI 仍可讓運算盒為更多攝影機執行智慧處理。
5. **機械視覺攝影機**。適用光學字元辨識、物件識別、瑕疵偵測和機械手臂導引的機械視覺攝影機運用嵌入式 AI 技術，進一步簡化產品開發並提升系統準確度。

表 1 列出各種應用程序的系統要求。

	ADAS	機器人	智慧型零售	機械視覺	邊緣 AI 運算盒
深度學習加速器	X	X	X	X	X
多攝影機影像訊號處理 (ISP)	X	X	X	X	X
視覺加速器	X	X	X	X	X
深度與運動加速器	X	X	X	X	X
乙太網路開關	X	X			X
快捷外設互聯標準 (PCIe) 交換器	X	X			
功能安全	X	X			

表 1. 邊緣 AI 系統的主要處理和元件需求。

什麼是高效邊緣 AI 系統？

在高效邊緣 AI 系統中，DNN 無法自行運作。高效 AI 系統需要複雜的視覺管線，通常包含單一或多個攝影機影像處理、傳統電腦視覺，甚至需要多個 DNN。有些應用可能還需影片編碼器和解碼器。為處理所有輸入，系統需要高性能運算。此外，系統可能需要強化安全性和功能安全，以增加系統複雜性和成本。

高效邊緣 AI 系統應針對下列項目進行最佳化：

- **性能**。嵌入式處理器必須能夠提供系統所需的速率、延遲和準確度，而且即使在嚴苛的環境中也能可靠地運作。
- **設計限制**。嵌入式處理器必須在有電源和散熱限制的設計中運作，包括無風扇設計、具有被動冷卻功能或需要以電池供電方式長時間運作的設計。處理器也必須符合尺寸與重量規格，以符合實體限制。

- **成本**。實現高性能且具成本效益的處理能力，進而獲得最低可物料清單 (BOM) 成本。

若要打造高效邊緣 AI 系統，設計人員應考慮哪些架構和核心最適合完成系統所需工作。

選擇 SoC 架構

嵌入式處理器設計選項有同質架構和異質架構兩種類型，通常整合專用處理功能來處理特定工作。您應根據所需的核心理型，評估最符合您邊緣 AI 系統需求的架構。

邊緣 AI 系統的目標是在最適合的核心上執行 AI、視覺、視訊和其他工作，讓系統在每瓦性能、每秒每 TOPS 性能、成本、尺寸和重量上都能提供最佳效果。擁有適合任務適合核心的異質架構，對邊緣 AI 系統而言至關重要。

並非所有具異質架構的處理器都採用相同設計。矽晶廠商必須選擇適當的處理功能或程序，並決定要對硬體中這些功能進行加速，或採用可配置或可編程方式處理。廠商也必須注意如何將核心整合到系統中。匯流排架構和記憶體子系統必須能夠在核心間有效率地移動資料。

如果 SoC 的工作加速核心理型不正確，或未有效率管理的核心數量過多，或匯流排基礎架構和記憶體子系統效率不佳，視覺邊緣 AI 系統就無法有效發揮作用。

可編程核心理型與加速器

我們來回顧一下邊緣 AI 系統中可能存在的核心理型：

CPU

中央處理單元 (CPU) 是可處理循序工作負載的通用處理單元。具備優異的編程靈活性，並可從現有的大型代碼庫中獲益。多數邊緣 AI 系統通常有兩到八個 CPU 核心，可用來管理平台和功能豐富的應用。但僅有 CPU 的系統不適合像素級成像、電腦視覺和卷積神經網路 (CNN) 處理等高度專業化的任務。CPU 的功耗也很高，但傳輸量是各種核心理型中最低的。若將單核心 CPU 系統搭配 AI 加速和影像處理等專用硬體區塊，即可滿足低成本應用的功率預算需求。

GPU

圖形處理單元 (GPU) 擁有成千上百個小型核心，非常適合平行處理工作。GPU 原本是為了執行一系列圖形操作而設計，在深度學習應用中十分常見，對 DNN 訓練特別實

用。其中一個主要缺點是由於核心數高，GPU 會消耗大量功率，晶片內建記憶體需求也較高。

DSP

數位訊號處理器 (DSP) 是一款具電源效率的專用核心，通常專為解決多個複雜數學問題而設計。DSP 可以低功耗處理真實世界視覺、音訊、語音、雷達和聲納感測器的即時資料。DSP 可幫助將每個時脈週期處理最大化。但在編程上並不容易，需要熟悉 DSP 硬體的功能、編程環境及 DSP 軟體最佳化，才能達到最佳性能。

ASIC

應用特定積體電路 (ASIC) 與加速器可以最低功耗為系統應用提供最大性能。若您了解欲加速功能的核心，此方式即是熱門選擇。例如 CNN 的核心運算一向涉及矩陣乘法。針對傳統電腦視覺工作，專用硬體加速器可運算影像縮放、鏡頭失真校正和雜訊濾波等作業。

FPGA

現場可編程邏輯陣列 (FPGA) 是一種積體電路，可針對特定應用重新編程並鎖定硬體區塊。此方式的功耗比 GPU 和 CPU 低，但使用的功率比 ASIC 多。但其硬體編程較不方便，且需對硬體描述語言 (如 Verilog 或超高速 IC 硬體描述語言) 的專業知識。

以 TI 視覺處理器設計邊緣 AI 系統

TI 的視覺處理器產品組合專為以尺寸與電源限制為主要設計挑戰的應用實現高效率、可擴充 AI 處理而打造

這些處理器包含 AM6xA 與 TDA4 處理器系列，具備含大量整合視覺系統的 SoC 架構，其中包括 Arm® Cortex®-A72 或 Cortex-A53 CPU、內部記憶體、介面和硬體加速器，可以每秒 2 至 32 兆次運算 (TOPS) 的 AI 處理能力進行深度學習。

AM6xA 系列使用 Arm Cortex-A MPU，將深度學習推論、成像、視覺、視訊與圖形處理等運算密集工作卸載到專用硬體加速器和可編程核心，如 圖 2 所示。在處理器中整合進階系統元件，可幫助邊緣 AI 設計人員簡化系統物料清單。此處理器產品組合包括以一至兩部攝影機進行低功耗應用的 AM62A 處理器可擴充處理選項，到 AM68A (最多八部攝影機) 和 AM69A (最多 12 部攝影機)。

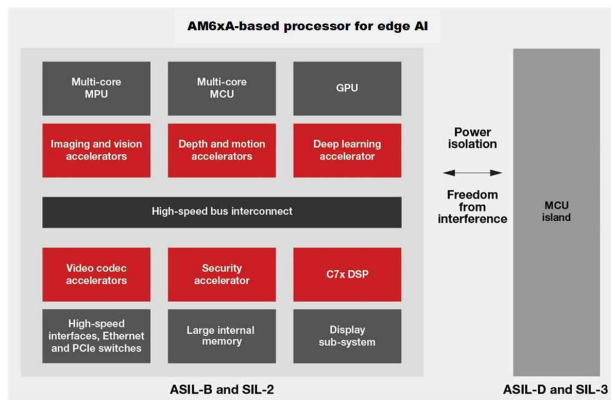


圖 2. TI 視覺處理器邊緣 AI 系統分割。

深度學習加速器

CPU 和 GPU 雖然適合其他工作，但不是加快深度學習工作的最佳核心。CPU 有輸送量限制且需高功耗；GPU 則是所有核心中消耗最多功率者，且記憶體體積龐大。

TI 視覺 AI 處理器整合深度學習加速器，其中包含栓接至可編程 C71 DSP，位於 ASIC 中的矩陣乘法加速器 (MMA)。MMA 可實現高性能 (每週期累積 4K 8 位元固定乘數) 和低功率張量加速，C71 DSP 則可加速向量和分頻器運算，並管理 MMA。

結合 MMA 與 C71 DSP 即可得到實現業界最高性能 (每秒推論) 與功率 (每瓦推論) 的加速器。C71 核心的編程靈活性讓您能夠跟上邊緣 AI 創新的腳步。除了應用在深度學習外，核心也可以低功耗處理其他運算密集工作。

智慧型記憶體架構可提高加速器的利用率。加速器有自己的記憶體子系統，適合資料傳輸的專用 4D 可編程直接記憶體存取 (DMA) 引擎，以及可將資料直接從外部記憶體傳送至 C71 核心和 MMA 功能單元並旁路快取的專用串流硬體。圖磚和超圖磚功能可減少來往外部記憶體間的資料傳輸。

表 2 說明採用每秒 8 TOPS 加速器的 AM68A 與 TDA4VM 8 位元固定推論性能。報告性能使用批次大小 1 及單一 32 位元 LPDD4。

網路	影像解析度	每秒幀數 (fps)
MobileNet_v1	224 × 224	741
Resnet-50 V1.5	224 × 224	162
SSD-MobileNets-V1	300 × 300	385

表 2. MLPerf 建議模型的推論測試基準。

免責聲明：TI 採用 MLPerf 建議模型與準則進行邊緣 AI 推論測試基準。TI 尚未將結果提交給 MLcommons 組織。

成像和電腦視覺硬體加速器

視覺型邊緣 AI 系統通常包含單一或多攝影機影像處理及傳統電腦視覺工作。在 CPU 或 GPU 中，這些工作會消耗大量功率，並有傳輸量限制。

此等級的邊緣 AI 處理器 SoC 可加快硬體中運算密集的低階暴力像素處理視覺工作，例如 ISP、鏡頭失真修正、多重調整及視覺處理加速器核心中的雙向雜訊濾波。深度和運動感知加速器核心可加快立體深度評估與密集光流，有助於增強對環境的認知，如 **圖 3** 中所示。



圖 3. 視覺加速器功能。

在硬體中加快這些工作可降低功耗並縮小體積。這些工作雖然在硬體中加速進行，但其可配置性透過加速器功能來滿足您的系統需求，為您提供更多靈活性。

此類整合與加速可省去對自訂 ISP 或 FPGA 的需求，並釋出 CPU Mega Hertz 以處理硬體中運算密集的成像及視覺工作。例如單一視覺處理加速器核心可在 30 fps 下處理高達八個 2 百萬像素或兩個 8 百萬像素的攝影機。深度和運動處理加速核心可以每秒 8 千萬像素進行立體深度評估，並以每秒 1 億 5 千萬像素進行動作向量。

智慧型內部匯流排與記憶體架構

監控處理器的資料移動和記憶體架構，避免同時執行多個核心時發生各種核心阻斷和延遲，可幫助發揮整體系統效能。

TI 視覺 AI 處理器具備與非阻斷架構與大型內部記憶體的高頻寬匯流排互連。多個專用可編程 DMA 引擎可在極高速度下將資料移動自動化。此設計可實現硬體加速器的高使用率，並可大幅節省雙數據速率 (DDR) 頻寬。減少 DDR 執行個體數量可降低 DDR 存取所用的功率量，進而降低整體系統功耗。

系統 BOM 最佳化

我們來回顧一下 TI 視覺 SoC 中的進階整合系統元件與功能，可為多種邊緣 AI 應用降低系統 BOM 成本：

- **ISP**。整合式 ISP 核心可省去對外部 ISP 或 FPGA 設計的需求。機械視覺、智慧型購物車、機器人和 ADAS 等所有單一和多攝影機 AI 應用都可因此整合獲益。
- **安全**。整合式汽車安全完整性等級 (ASIL) D 與 SIL 3 相容安全微控制器 (MCU) 與 Cortex-R5 核心，可幫助在沒有外部安全 MCU 的情況下實現安全目標。此架構的其他處理作業也符合 ASIL B/SIL 2 規範，故可實現 ADAS、機器人、建築與農業電子控制單元應用。
- **乙太網路和 PCIe 交換器**。整合式乙太網路與 PCIe 交換器可省去對外部交換器元件的需求。
- **保全**。整合式安全加速器可提供最先進的安全支援。
- **DDR 記憶體**。相較於一般記憶體架構，內嵌式錯誤修正程式碼保護與較少 DDR 記憶體執行個體 (透過智慧型記憶體) 可節省成本。

使用簡單的軟體開發環境

圖 4 所示為 TI 的全方位軟體環境，讓您無需學習 TI 硬體或專有軟體，即可部署異質架構並獲得矽性能的最大潛能。透過生產品質驅動器萃取硬體加速器，並利用業界標準應用程式編程介面 (API) 為應用開發 MPU 上的高階作業系統提供介面，以實現更快速的軟體開發。TI 的低階軟體可自動加速成像、視覺、深度學習與多媒體工作至正確的硬體加速器，讓高效能應用程式編程變得更簡單。

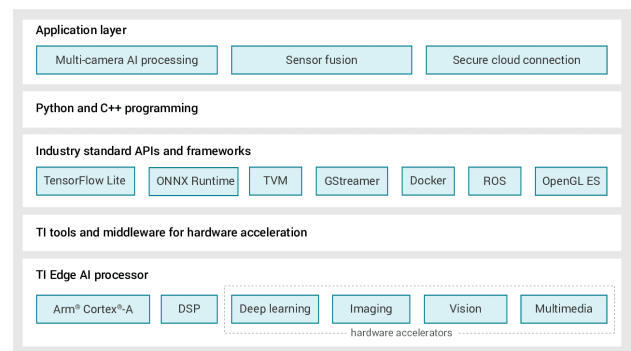


圖 4. 適用邊緣 AI 應用的軟體開發環境。

結論

異構架構在應用程序中的應用不斷增加。TI 的視覺 AI 處理器具備加速深度學習、視覺與視訊處理，以及專用系統整合與進階元件整合，讓市售可用邊緣 AI 系統能針對性能、電源、尺寸、重量和系統成本進行最佳化。TI 的邊緣 AI 軟體開發環境以開放原始碼業界標準 API 為基礎，並具備硬體加速器的自動加速功能，以加快邊緣 AI 應用開發。

AI 是快速演進的技術，可促進邊緣 AI 應用各方面創新。此技術也讓需要更高運算需求的應用不斷突破。透過執行嵌入式處理器在低功耗與低系統成本下啟用邊緣 AI 技術，可為嵌入式應用開創無限可能。

重要聲明：本文所述德州儀器及其子公司相關產品與服務經根據 TI 標準銷售條款及條件。建議客戶在開出訂單前先取得 TI 產品及服務的最新完整資訊。TI 不負責應用協助、客戶的應用或產品設計、軟體效能或侵害專利等問題。其他任何公司產品或服務的相關發佈資訊不構成 TI 認可、保證或同意等表示。

Arm® and Cortex® are registered trademarks of Arm Limited (or its subsidiaries) in the US and/or elsewhere.
所有商标均为其各自所有者的财产。

IMPORTANT NOTICE AND DISCLAIMER

TI PROVIDES TECHNICAL AND RELIABILITY DATA (INCLUDING DATA SHEETS), DESIGN RESOURCES (INCLUDING REFERENCE DESIGNS), APPLICATION OR OTHER DESIGN ADVICE, WEB TOOLS, SAFETY INFORMATION, AND OTHER RESOURCES "AS IS" AND WITH ALL FAULTS, AND DISCLAIMS ALL WARRANTIES, EXPRESS AND IMPLIED, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT OF THIRD PARTY INTELLECTUAL PROPERTY RIGHTS.

These resources are intended for skilled developers designing with TI products. You are solely responsible for (1) selecting the appropriate TI products for your application, (2) designing, validating and testing your application, and (3) ensuring your application meets applicable standards, and any other safety, security, regulatory or other requirements.

These resources are subject to change without notice. TI grants you permission to use these resources only for development of an application that uses the TI products described in the resource. Other reproduction and display of these resources is prohibited. No license is granted to any other TI intellectual property right or to any third party intellectual property right. TI disclaims responsibility for, and you will fully indemnify TI and its representatives against, any claims, damages, costs, losses, and liabilities arising out of your use of these resources.

TI's products are provided subject to [TI's Terms of Sale](#) or other applicable terms available either on [ti.com](https://www.ti.com) or provided in conjunction with such TI products. TI's provision of these resources does not expand or otherwise alter TI's applicable warranties or warranty disclaimers for TI products.

TI objects to and rejects any additional or different terms you may have proposed.

Mailing Address: Texas Instruments, Post Office Box 655303, Dallas, Texas 75265
Copyright © 2023, Texas Instruments Incorporated