

使用 FP-AI-VISION1 的影像分類器

作者：意法半導體

1. ST AI 解決方案簡介

意法半導體身為全球最大的半導體公司之一，持續積極參與高速發展的嵌入式人工智慧領域。為了在具成本效益和低功耗的微控制器上加快運用機器學習和深度神經網路，ST開發全方位的邊緣人工智慧生態系統，嵌入式開發人員可以在各種STM32微控制器產品組合中，輕鬆新增利用人工智慧的新功能和強大的解決方案。圖 1 顯示ST AI解決方案之於整個STM32產品組合，而且已經擁有預先訓練神經網路的嵌入式開發人員，可以在任何採用Cortex M4、M33和M7的STM32上移植、最佳化和驗證這整個產品組合。STM32Cube.AI是 STM32CubeMX的AI擴充套件，讓客戶能以更高地效率開發其AI產品。

您可以利用深度學習的強大功能增強訊號處理效能，並提升STM32 應用的工作效率。

本文概述FP-AI-VISION1，這是用於電腦視覺開發的架構，提供工程師在STM32H7上執行視覺應用的程式碼範例。

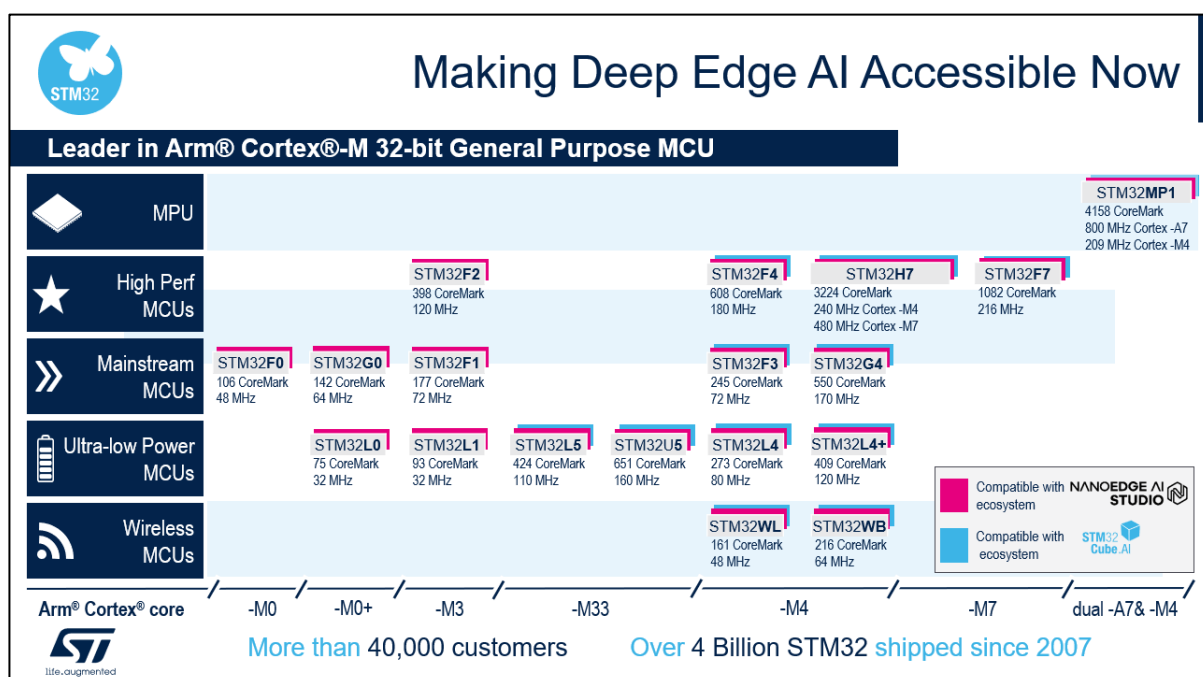


圖 1、執行機器學習和深度學習演算法的STM32產品組合

從FP-AI-VISION1程式碼範例開始，簡易達成在邊緣執行的各種電腦視覺使用案例，例如

- 生產線上的物體分類，藉以根據物體的類型調整輸送帶速度；
- 偵測產品的一般瑕疵；
- 區分不同類型的物件，例如：螺絲、義大利麵、樂高零件，並分揀到不同的容器中；
- 對設備或機器人操作的材料類型進行分類，並隨之調整行為；
- 對食品類型進行分類以調整烹調／燒烤／沖泡或重新訂購貨架上的新產品。

2. FP-AI-VISION1

2.1 概述

FP-AI-VISION1是STM32Cube功能套件（FP），提供採用卷積神經網路（CNN）的電腦視覺應用範例。其由STM32Cube.AI產生的軟體元件和AI電腦視覺應用專用的應用軟體元件組成。

功能套件中提供的應用範例包含：

- 食品識別：辨識 18 類常見食品；
- 人體感測：識別影像中是否有人；
- 人數統計：依照物體偵測模型計算情境中的人數。

2.2 主要特點

FP-AI-VISION1在與STM32F4DIS-CAM攝影機子板，或是B-CAMS-OMV攝影機模組搭配連接的STM32H747I-DISCO板上運作，包括用於攝影機擷取、畫格影像預處理、推斷執行的完整應用軟體和輸出後處理。這也提供浮點和 8 位元量化 C 模型的整合範例，並支援多種資料記憶體設定，滿足各種應用需求。

此功能套件最重要的其中一項關鍵優勢是提供範例，描述如何將不同類型的資料有效地放置在晶片上的記憶體和外部記憶體中。使用者因此能夠輕鬆瞭解最符合需求的記憶體分配，並有助建立適用於STM32系列的自訂神經網路模型，特別是在STM32H747-DISCO板上。



圖2、FP-AI-VISION1評估裝置範例

FP-AI-VISION1 包括三個採用 CNN 的影像分類應用範例：

- 一種對彩色（RGB 24 位元）畫格影像執行的食品識別應用；
- 一種對彩色（RGB 24 位元）畫格影像執行的人體感測應用；
- 一種對灰階（8 位元）畫格影像執行的人體感測應用。

本文將重點介紹食品識別和人體感測兩種範例。

首先討論食品識別應用。

食品識別CNN 是MobileNet模型的衍生模型。MobileNet是適用於行動和嵌入式視覺應用的高效率模型架構，此模型架構由 Google®[1]提出。

MobileNet模型架構包括兩個簡單的全域超參數，可以高效在延遲和準確度之間進行權衡。原則上，這些超參數可讓模型建構者根據問題的限制條件，決定應用大小適合的模型。考量 STM32H747的目標限制條件，此軟體套件中使用的食品識別模型是透過調整這些超參數建構而成，以便在準確度、運算成本和記憶體佔用之間進行最佳權衡。

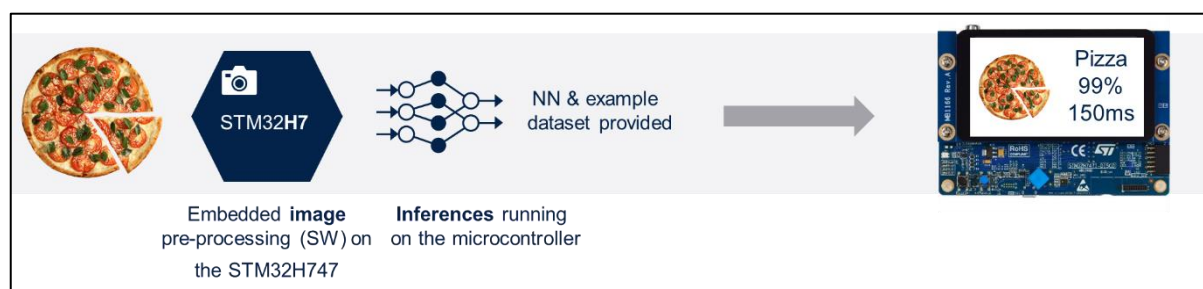



圖3、食品識別模型的執行流程

圖3為食品識別模型的簡單執行流程。這在STM32H747上執行，大約需要150毫秒才能完成推斷。其次，來說明人體感測應用。FP-AI-VISION1 提供兩個人體感測的範例應用：

- 一種採用低複雜度 CNN 模型（所謂的 Google_Model），用於處理解析度為 96×96 像素的灰階影像（每像素 8 位元）。這個模型可從：storage.googleapis.com 下載。
- 一種採用更高複雜度的 CNN 模型（所謂的 MobileNetv2_Model），用於處理解析度為 128×128 像素的彩色影像（每像素 24 位元）。

我們在這裡看看前一個模型。人體感測應用程式可識別影像中是否有人。在與STM32F4DIS-CAM連接的STM32L4R上執行這個應用程式時，大約需要270毫秒來推斷。而快閃記憶體和RAM 的大小足以在微控制器上執行神經網路 (NN)，如圖 4 所示。



Demo overview & features

- TensorFlowLite model converted to STM32 code using STM32Cube.AI
- Computer Vision - Image Classification integration example on the [STM32L4R + STM32F4DIS-CAM](#)

Person Detection model from [tensorflow.org](https://www.tensorflow.org)
Input shape: (96x96x1)
Output shape: (3,)

Inference time	274 ms @ 120 MHz
Flash	214 KB (NN weights)
RAM	46 KB (NN activations & input)
Complexity	69k MACC
Max rate	3.6 fps

圖4、人體感測概覽

人體感測可利用低功率喚醒，使用案例包括開燈、開門或其他任何自訂方法。一般用途通常採用被動紅外線感測器，藉以在偵測到移動的時間和地點觸發事件。不過，這種PIR 系統的問題是可能發生誤報。如果有貓經過或在風中飛舞的樹葉，可能會觸發這個系統。人體感測應用只會偵測人類，並且有助於輕鬆開發更智慧的偵測系統。

2.3 系統架構

FP-AI-VISION1的頂層架構如圖5所示。

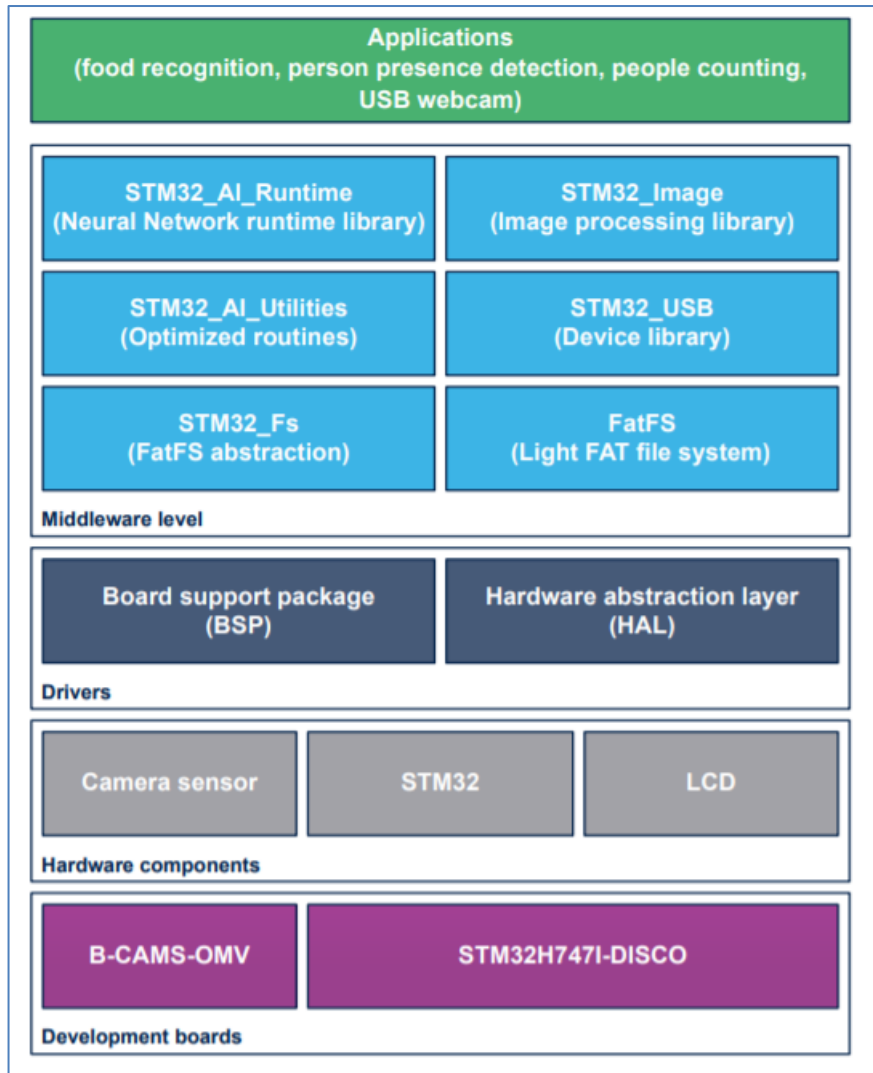


圖5、FP-AI-VISION1架構

2.4 應用建構流程

從浮點 CNN 模型（使用 Keras 等架構設計和訓練）開始，使用者產生最佳化的C程式碼（使用 STM32Cube.AI 工具）並整合到電腦視覺架構中（FP-AI-VISION1 提供），以便在 STM32H7 上建構電腦視覺應用。

產生C程式碼時，使用者可從下列兩個選項中擇一：

- 直接從CNN模型以浮點方式產生浮點C程式碼；
- 或者對浮點CNN模型進行量化，得到8位元模型，隨後產生對應之量化後的C程式碼；

對於大多數 CNN 模型，第二個選項可以減少記憶體佔用（快閃記憶體和RAM）以及推斷時間。對最終輸出準確度的影響則取決於CNN模型以及量化過程（主要是測試資料集和量化演算法）。

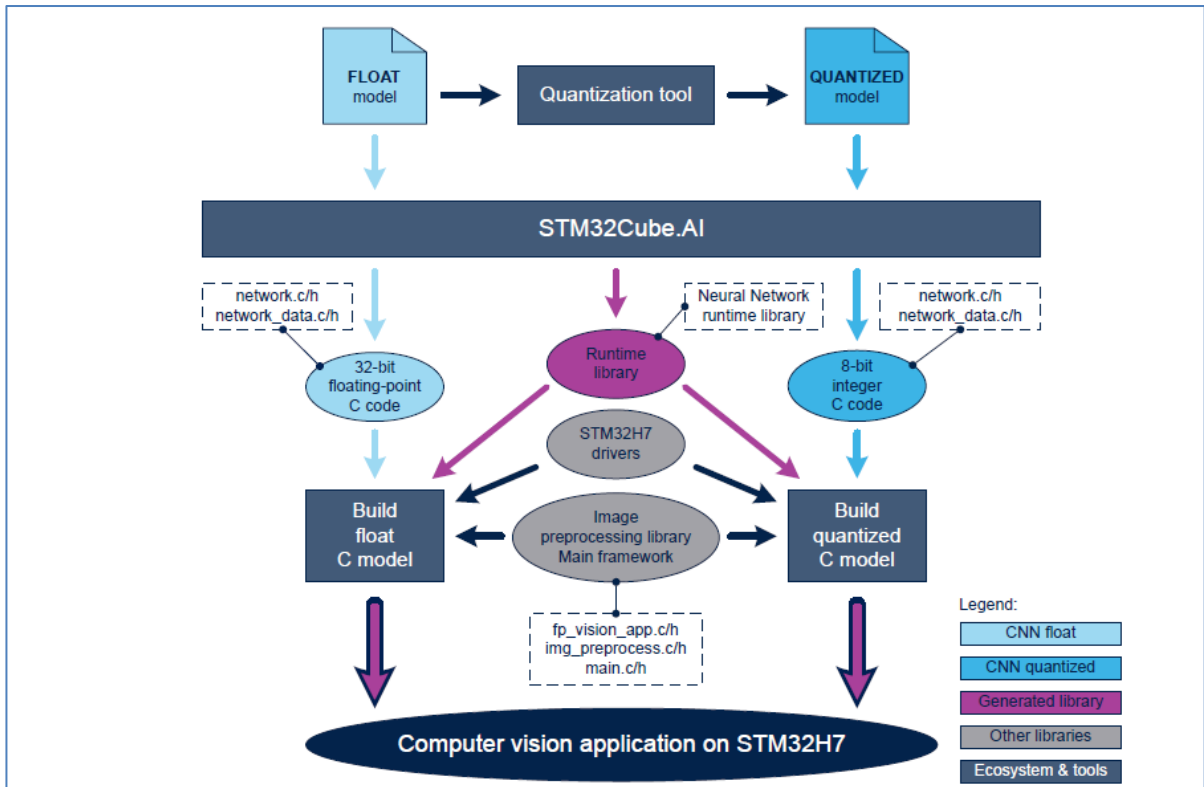


圖6、FP-AI-VISION1架構

2.5 應用執行流程

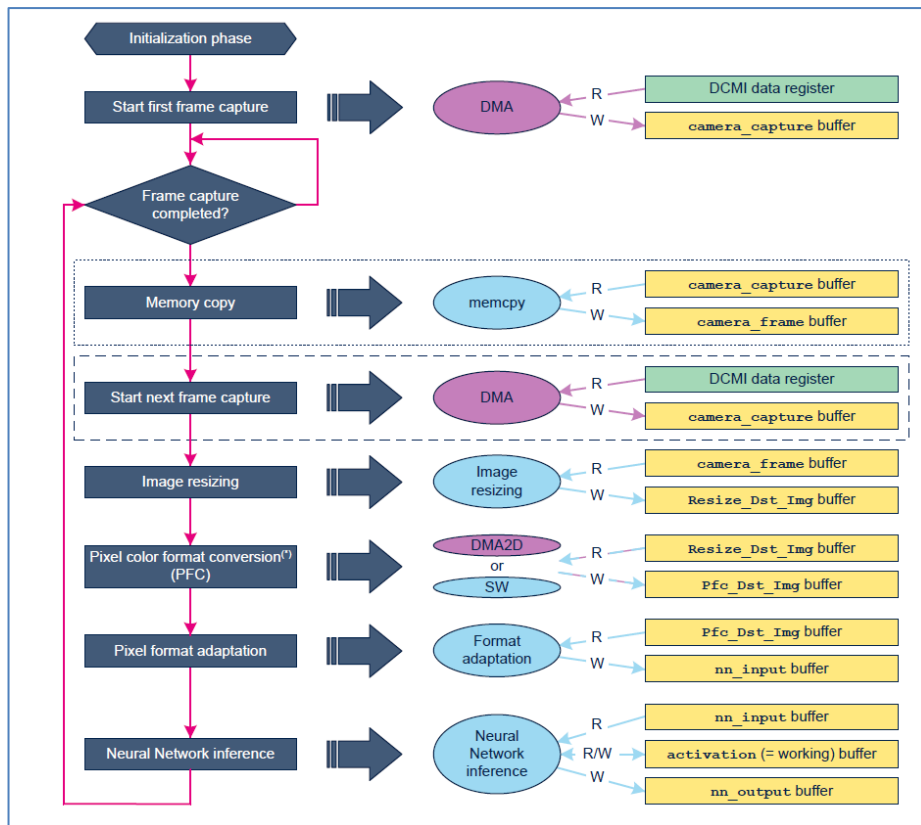


圖7、執行流程中的資料緩衝區

對於電腦視覺應用，整合需要一些資料緩衝區，如圖 7 所示。

應用過程依序執行下列操作：

1. 在 camera_capture buffer 中擷取攝影機畫格（透過 DCMI 資料暫存器中的 DMA 引擎）。
2. 此時，根據選擇的記憶體分配配置，將 camera_capture buffer 內容複製到 camera_frame buffer，並啟動後續畫格的擷取。
3. camera_frame buffer 中包含的影像將重新縮放到 Resize_Dst_Img buffer 中，藉以配合預期的 CNN 輸入張量尺寸。例如，食品識別 NN 模型需要輸入張量，例如 Height × Width = 224 × 224 像素。
4. 執行 Resize_Dst_Img buffer 到 Pfc_Dst_Img buffer 的像素色彩格式轉換。
5. 將 Pfc_Dst_Img buffer 內容中包含的各像素格式調整到 nn_input 緩衝區中。
6. 執行 NN 模型的推斷：這個 nn_input buffer 以及 activation buffer 提供給 NN 作為輸入。分類結果將儲存在 nn_output buffer 中。
7. 對 nn_output buffer 內容進行後處理，並顯示結果在顯示器上。

下方表 1 詳細說明食品識別應用在整合量化 C 模型或浮點 C 模型時所需的 RAM。

表1、用於食品識別應用的SRAM記憶體緩衝器

SRAM data buffer		SRAM data buffer size (byte)			
		Food recognition CNN			
Name	Pixel format	Quantized C model		Float C model	
		VGA capture (640 × 480)	QVGA capture (320 × 240)	VGA capture (640 × 480)	QVGA capture (320 × 240)
camera_capture	16 bits (RGB565)	600 K (640 × 480 × 2)	150 K (320 × 240 × 2)	600 K (640 × 480 × 2)	150 K (320 × 240 × 2)
camera_frame	16 bits (RGB565)	600 K (640 × 480 × 2)	150 K (320 × 240 × 2)	600 K (640 × 480 × 2)	150 K (320 × 240 × 2)
Resize_Dst_Img	16 bits (RGB565)	98 K (224 × 224 × 2)			
Pfc_Dst_Img	24 bits (RGB888)	147 K (224 × 224 × 3)			
nn_input ⁽¹⁾	24 bits (RGB888)	147 K (224 × 224 × 3)		588 K (224 × 224 × 3 × 4)	
activation ⁽¹⁾	-	98 K		395 K	
nn_output	-	72 (18 × 4)		72 (18 × 4)	

3. 想要進一步瞭解？

FP-AI-VISION1可從連結下載：

www.st.com/en/embedded-software/fp-ai-vision1.html。

欲深入瞭解意法半導體微控制器 AI 解決方案，請瀏覽下方資源網站：

UM2611：FP-AI-VISION1手冊

https://www.st.com/resource/en/user_manual/dm00630755-artificial-intelligence-ai-and-computer-vision-function-pack-for-stm32h7-microcontrollers-stmicroelectronics.pdf

UM2526：STM32Cube.AI手冊

https://www.st.com/resource/en/user_manual/dm00570145-getting-started-with-xcubeai-expansion-package-for-artificial-intelligence-ai-stmicroelectronics.pdf

FP-AI-VISION1 視訊應用影片：

<https://www.youtube.com/watch?v=VkFTrc5KSgg&t=130s>

<https://www.youtube.com/watch?v=8AX9uC2Oi1g&t=178s>

4. 參考資料

[1] 食品識別 CNN 是 MobileNet 模型的衍生模型。

MobileNet：用於行動視覺應用的高效率卷積神經網路：

<https://arxiv.org/pdf/1704.04861.pdf>