

AN14132

在MCX N947上利用神经处理单元（NPU）加速人脸检测演示

第1版 — 2024年1月20日

应用笔记

文档信息

信息	内容
关键词	AN14132、MCX N94x、NPU、机器学习
摘要	本应用笔记演示如何在FRDM-MCXN947板上实现人脸检测示例。



1 介绍

本应用笔记演示如何在FRDM-MCXN947板上实现人脸检测示例。

MCX N94x和MCX N54x基于两个运行频率高达150 MHz的高性能Arm Cortex-M33核，配备2 MB的Flash和可选的完整ECC RAM，并集成了专有神经处理单元（NPU）。与CPU核相比，集成的NPU机器学习的推理性能提升40倍，极大的减少运行时间，从而降低整机功耗。

表1. 人脸检测要求

软件要求	• MCUXpresso IDE • MCUXpresso IDE • eIQ工具包 • MCUXpresso SDK for FRDM-MCXN947 • MCUXpresso SDK for FRDM-MCXN947
硬件要求	• FRDM-MCXN947板 • LCD显示屏：LCD-PAR-S035 • 摄像头模块：OV7670 • 调试探头：USB-C

2 NPU概述

eIQ NPU是一种高度可扩展的加速器核心架构，提供ML加速。该架构提供功率和性能优化的NPU的支持，并与恩智浦的广泛微控制器和应用处理器产品组合相集成。

eIQ NPU支持多种类型的神经网络，如CNN、RNN、TCN和Transformer网络等。eIQ ML软件开发环境集成了NPU驱动及应用示例。

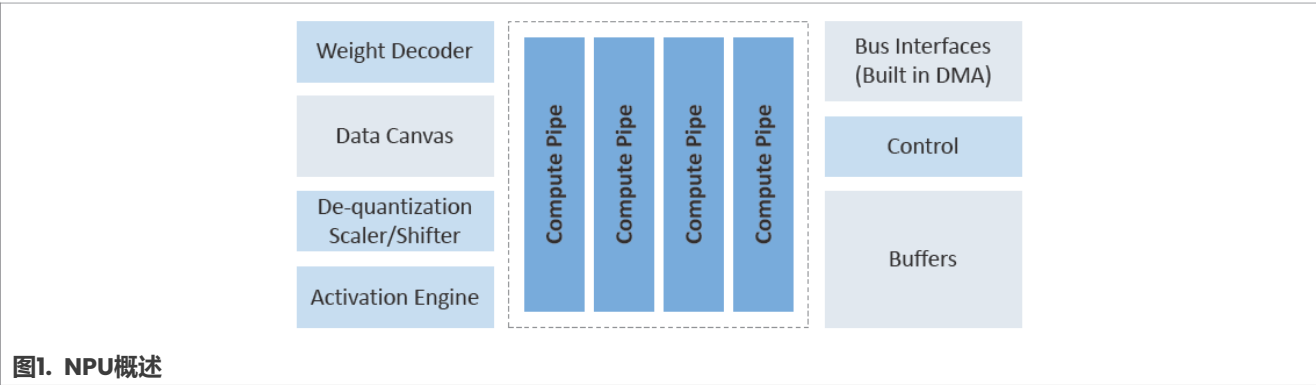


图1. NPU概述

MCX N94中的eIQ Neutron N-16拥有四个计算管道（compute pipe），每个计算管道包含四个MAC（乘积累加运算计算器），因此共有16个MAC。这意味着MCX N94每秒可以执行4.8 G（150 MHz × 4 × 4 × 2）次运算。

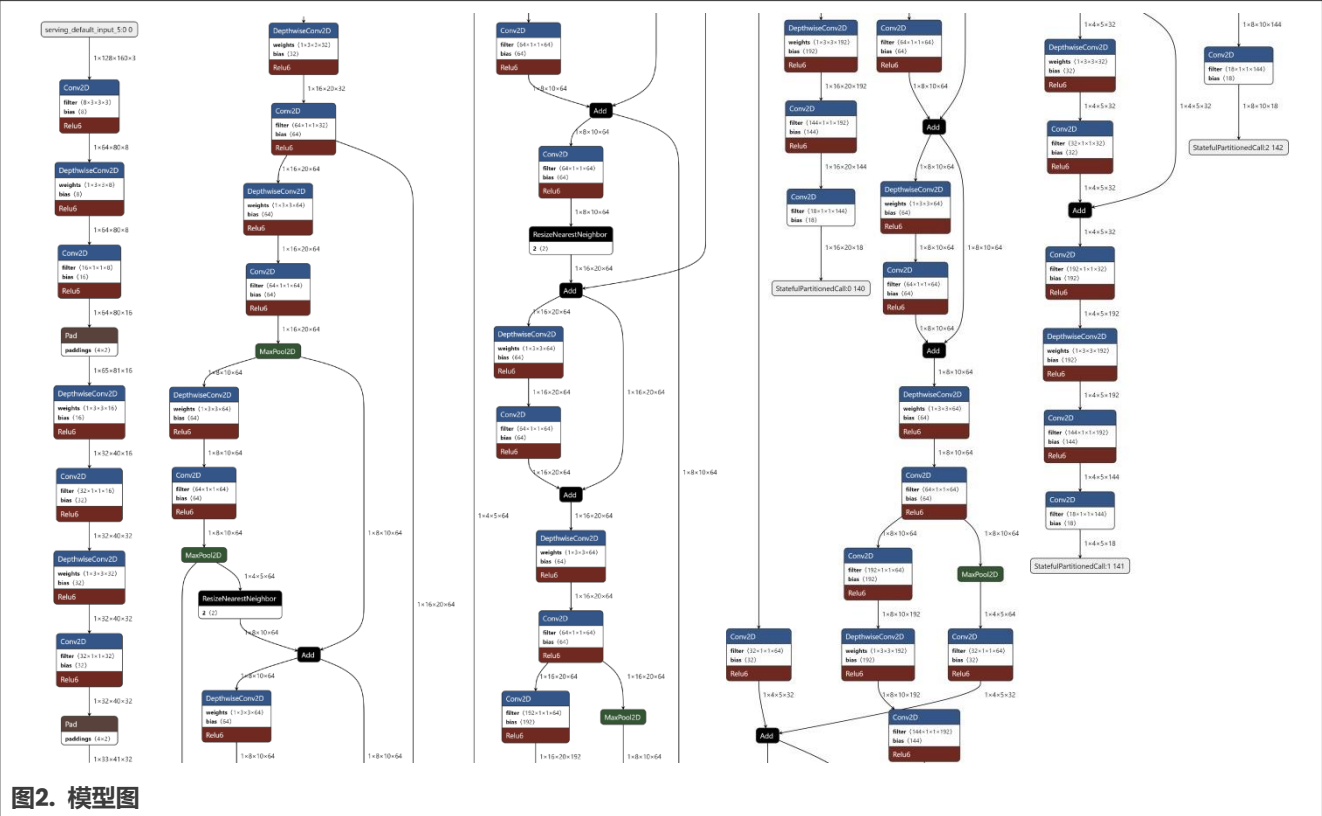
eIQ Neutron支持常见的神经网络运算符，如Conv2D、Depthwise Conv2D、full connected、Add、Average Pooling2D、Maximum Pooling2D和Padding,它还支持8位量化权重。

3 人脸检测模型

本演示运行一个YOLO模型，以检测摄像头捕获的帧中的多个人脸位置。这个YOLO模型经过了优化，只有原始YOLOV3大小的1/36。

为了适应摄像头帧大小，将模型输入大小设置为160 * 128。该模型有三路输出，可以检测三种大小的人脸：大尺寸人脸、中尺寸人脸和小尺寸人脸。

图2显示了模型图。



为了使用eIQ Neutron加速此模型，请首先使用eIQ工具包中的模型转换工具：Neutron Converter转换该模型。

如图3和图4所示，完成以下步骤：

- 1. 从[EIQ-TOOLKIT](#)中下载eIQ工具包安装包。
- 2. 双击下载的安装包。
- 3. 选择安装位置。
- 4. 进行安装，直至完成。

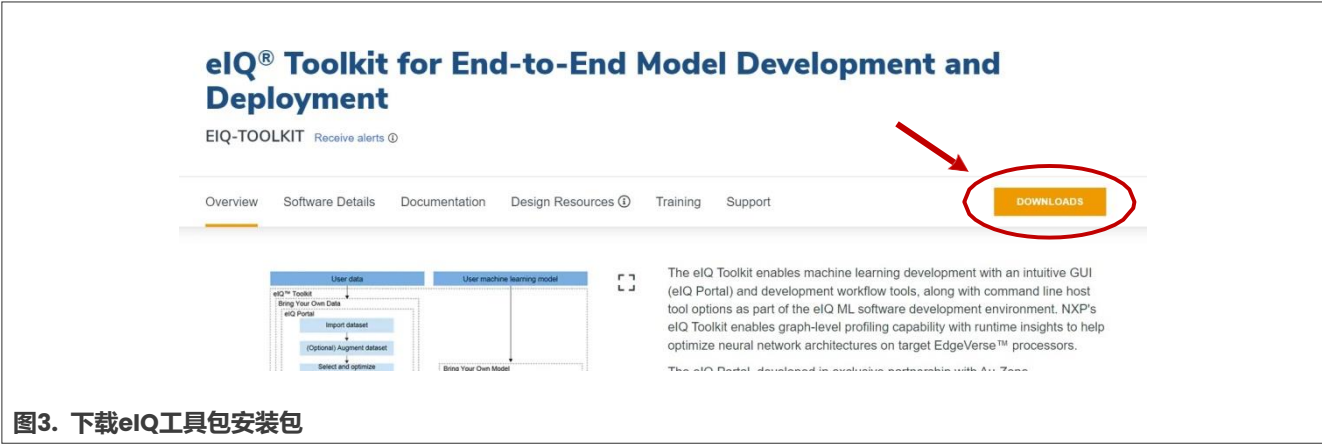


图3. 下载eIQ工具包安装包

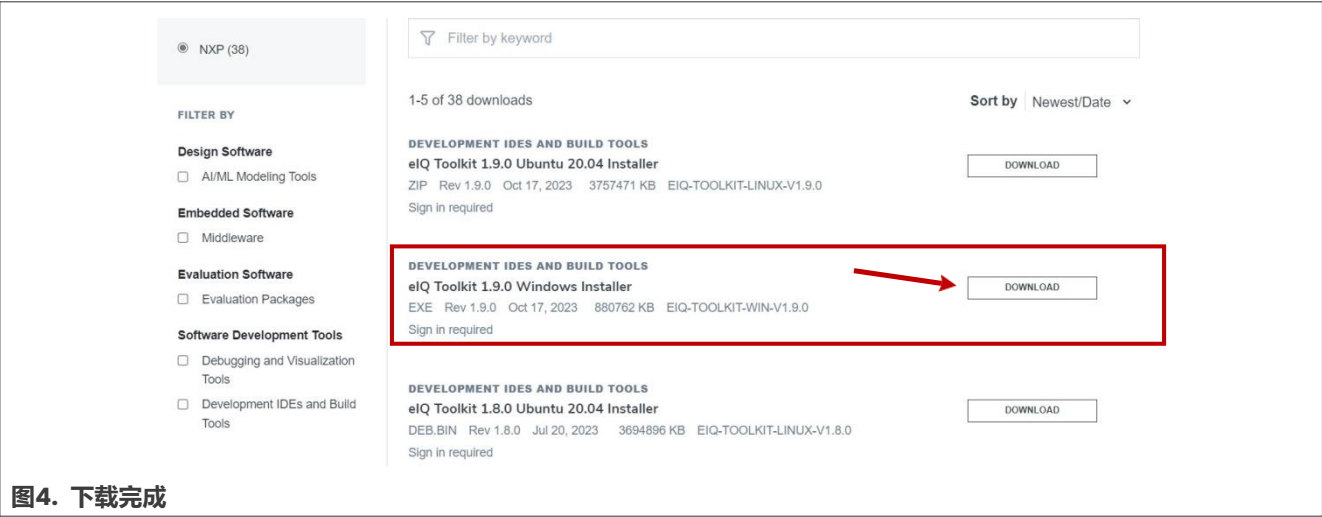


图4. 下载完成

打开命令行并进入eIQ工具包文件夹：C:\npx\eiQ_Toolkit_v2.10.1\bin\neutron-converter\v1.2.0>。将模型文件复制到该文件夹中。执行命令，转换模型，如图5所示。

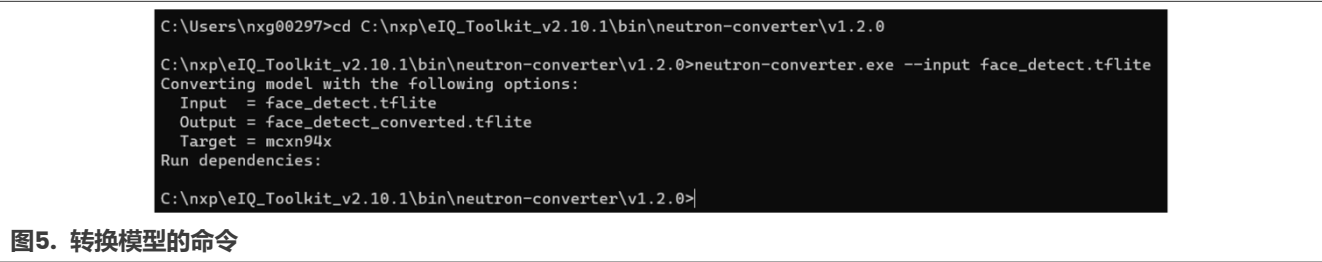


图5. 转换模型的命令

模型转换器工具更改模型的计算图，将在NPU中运行的运算符转换为自定义的NeutronGraph节点，如图6所示。转换后的模型命名为face_detect_converted.tflite。

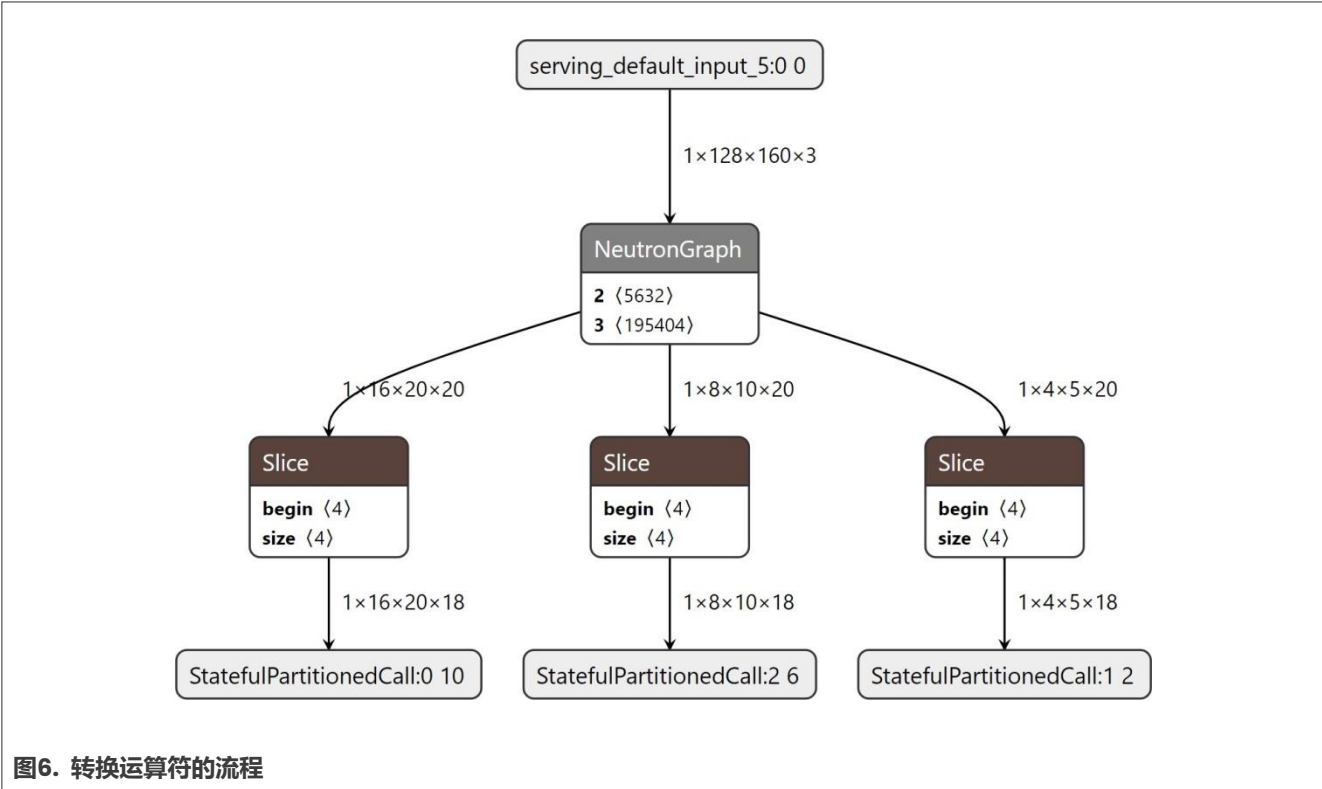


图6. 转换运算符的流程

转换后，运算符数量变少，量化因子的表达更加高效。[表2](#)对转换前后的模型进行了对比。

表2. 转换前后的模型对比

	原始模型	转换后的模型
文件大小	287 kB	199 kB
张量区域 (Tensor arena)	194 kB	208 kB

4 硬件和软件环境设置

如图7所示，查看下方红色和黄色框内的连接位置。连接摄像头，并将LCD模块的引脚与接头（header）的尾部对齐。在这里，选择MCU-Link进行下载和调试，其连接方式如浅蓝色方框所示。此外，使用USB数据线来为MCX-N947板供电。

LCD显示器LCD-PAR-S035是一个分辨率为320×480、大小为3.5英寸的显示屏。LCD模块可以通过恩智浦官网购买。

摄像头模块为OV7670，像素为640×480，支持最大为FPS 30的VGA。摄像头也可以通过亚马逊、淘宝网等在线购物平台购买。

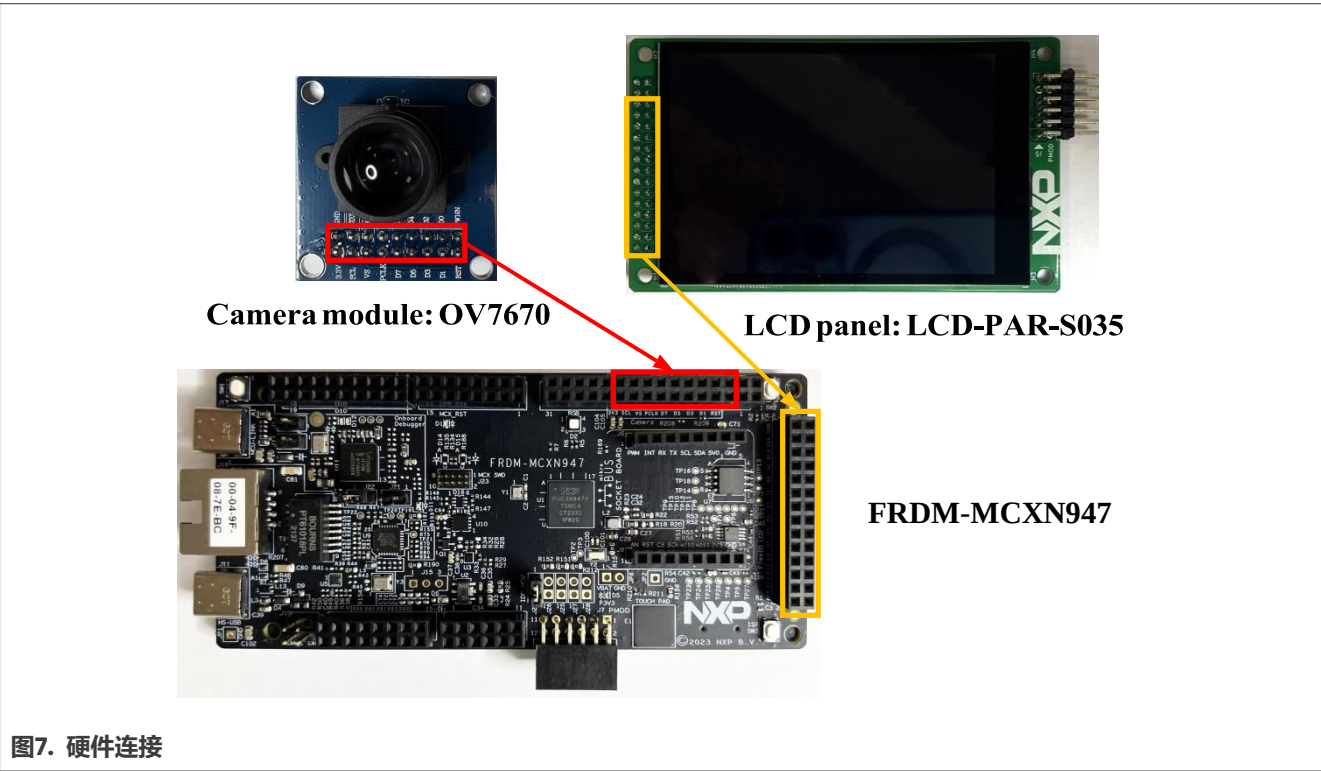
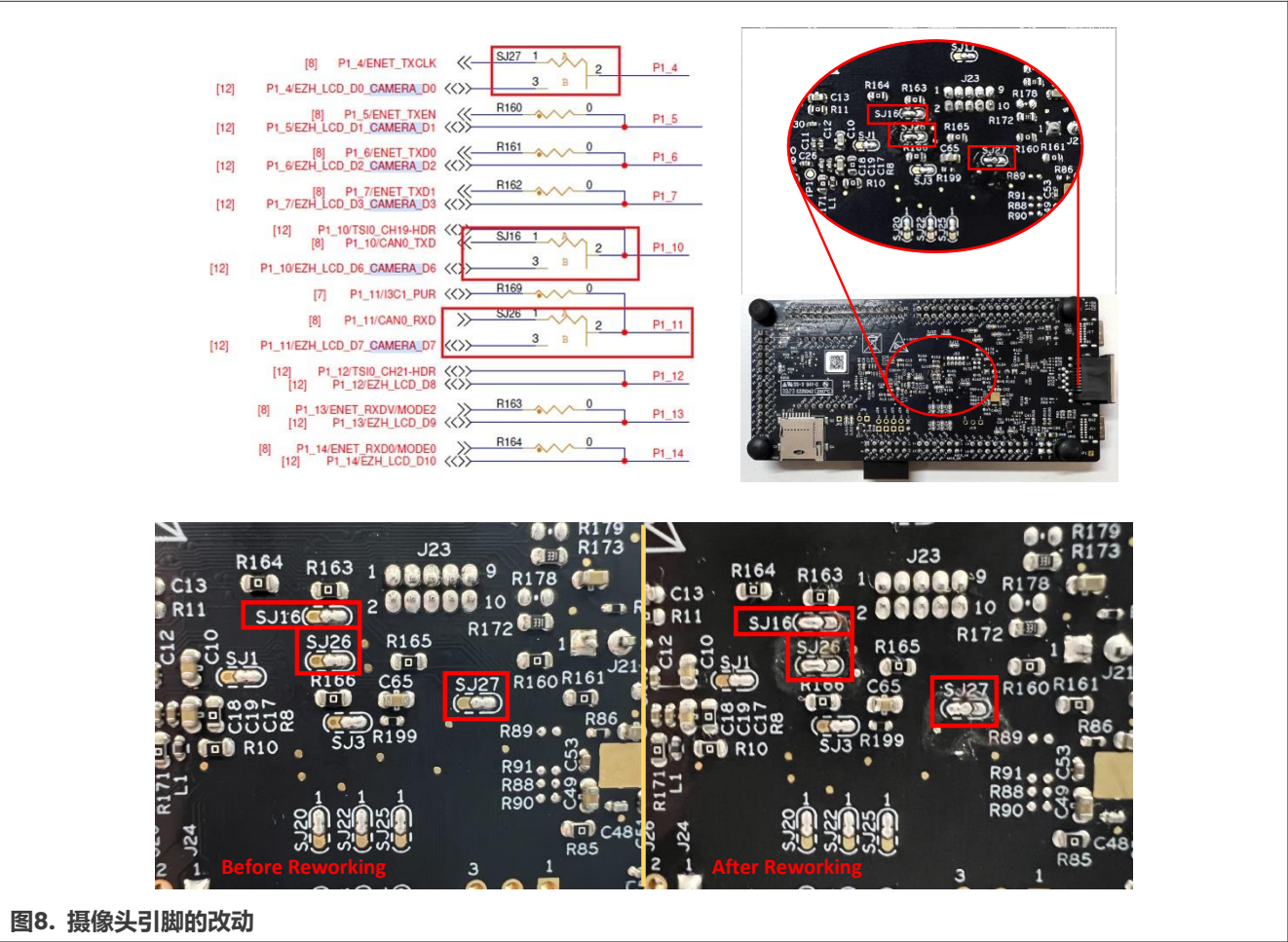


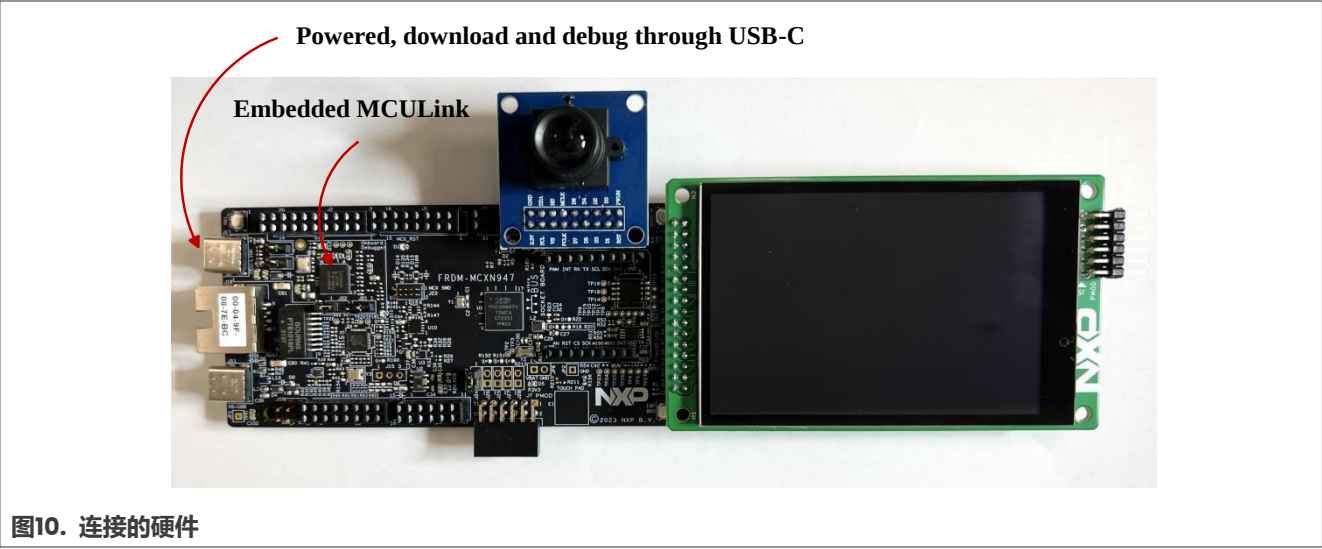
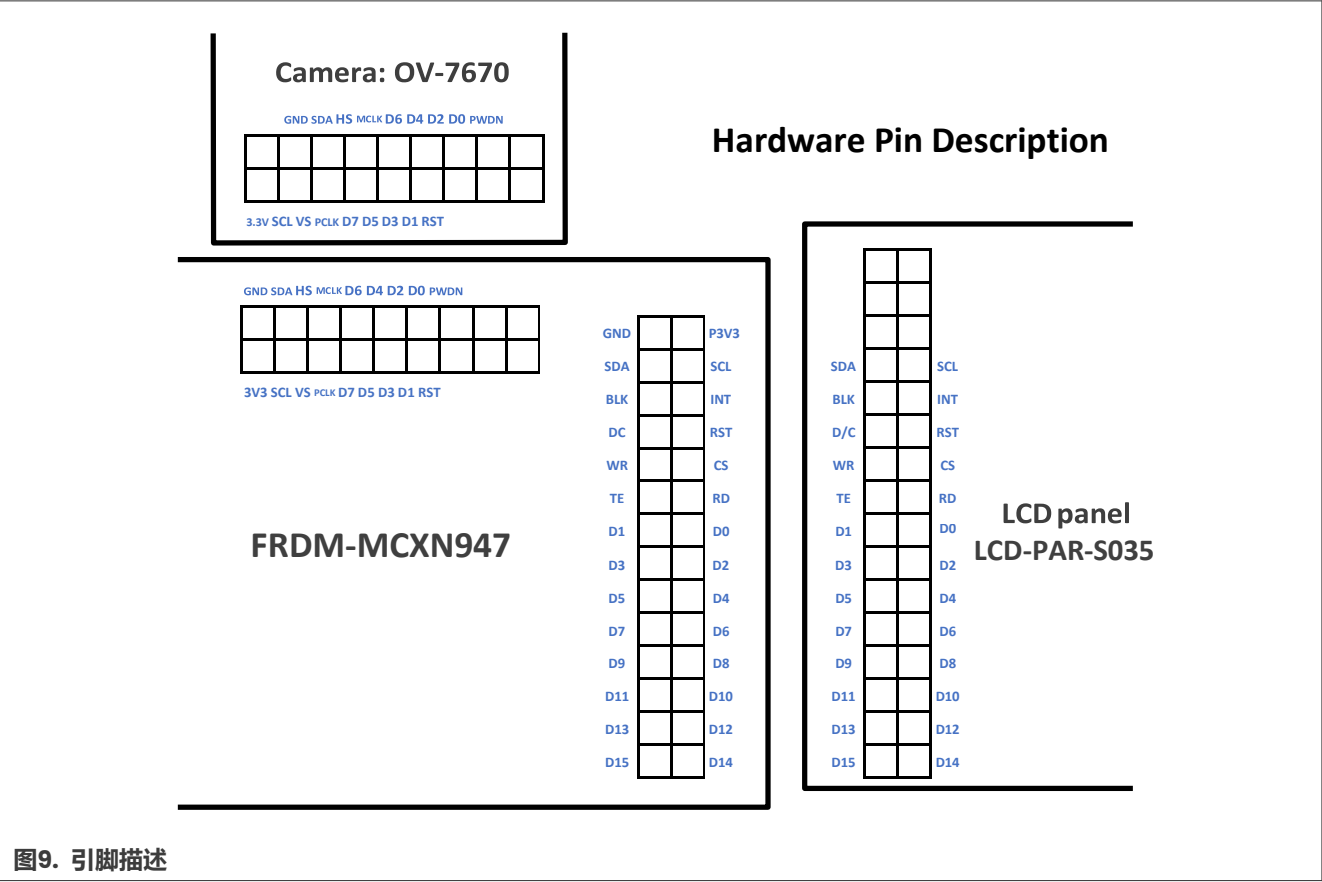
图7. 硬件连接

在连接硬件之前，我们需要对FRDM-MCXN947上的摄像头引脚进行一些改动，因为使用默认配置摄像头工作异常。如图8所示，将SJ16、SJ26、SJ27从A侧改到B侧。

在MCX N947上利用神经处理单元（NPU）加速人脸检测演示



完成上述调整后，硬件准备就绪，可以组装在一起。图9展示了引脚，图10展示了连接的硬件，留出两个USB口用于供电和程序下载及调试。



软件环境的设置基于Windows 11。下载[MCUXpresso IDE](#)并选择下载Windows版本，如[图11](#)所示。

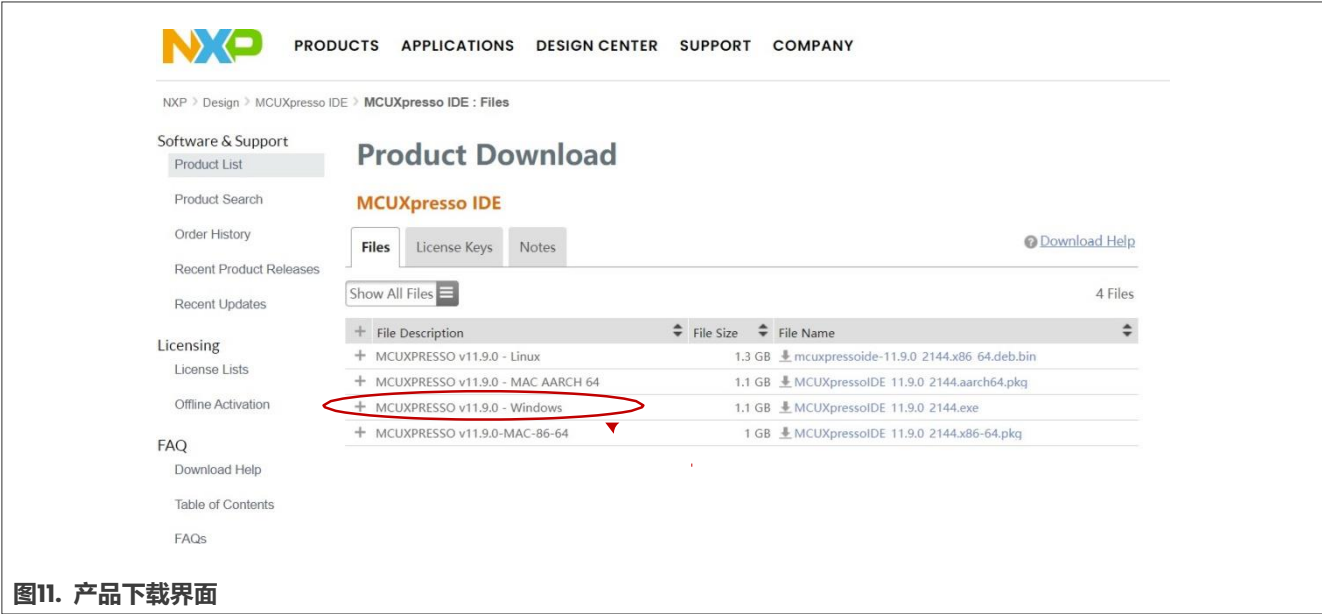


图11. 产品下载界面

双击下载的安装包，选择安装位置，并执行安装，直到完成。

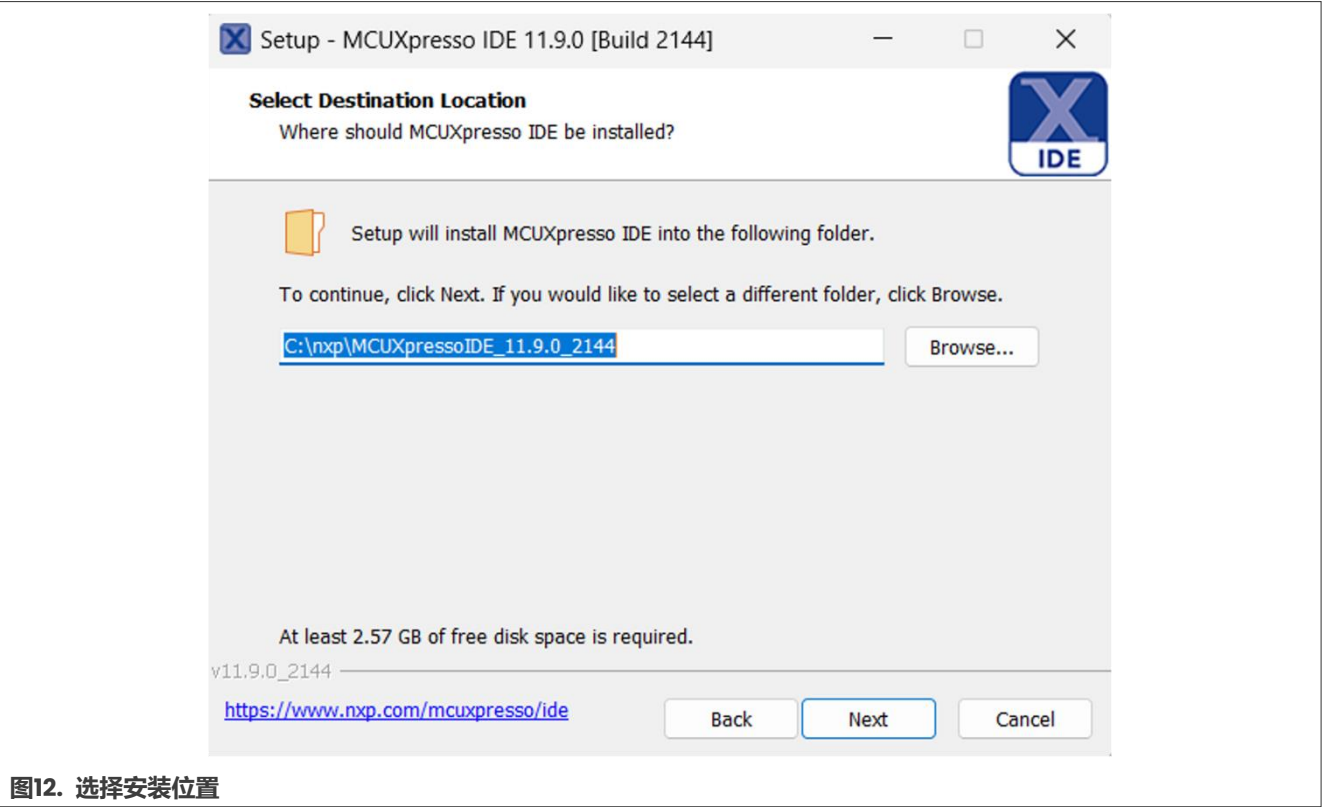


图12. 选择安装位置

使用一根USB-C数据线来为FRDM-MCXN947板供电和进行调试，并将另一端连接到计算机，如图13所示。

在MCX N947上利用神经处理单元（NPU）加速人脸检测演示

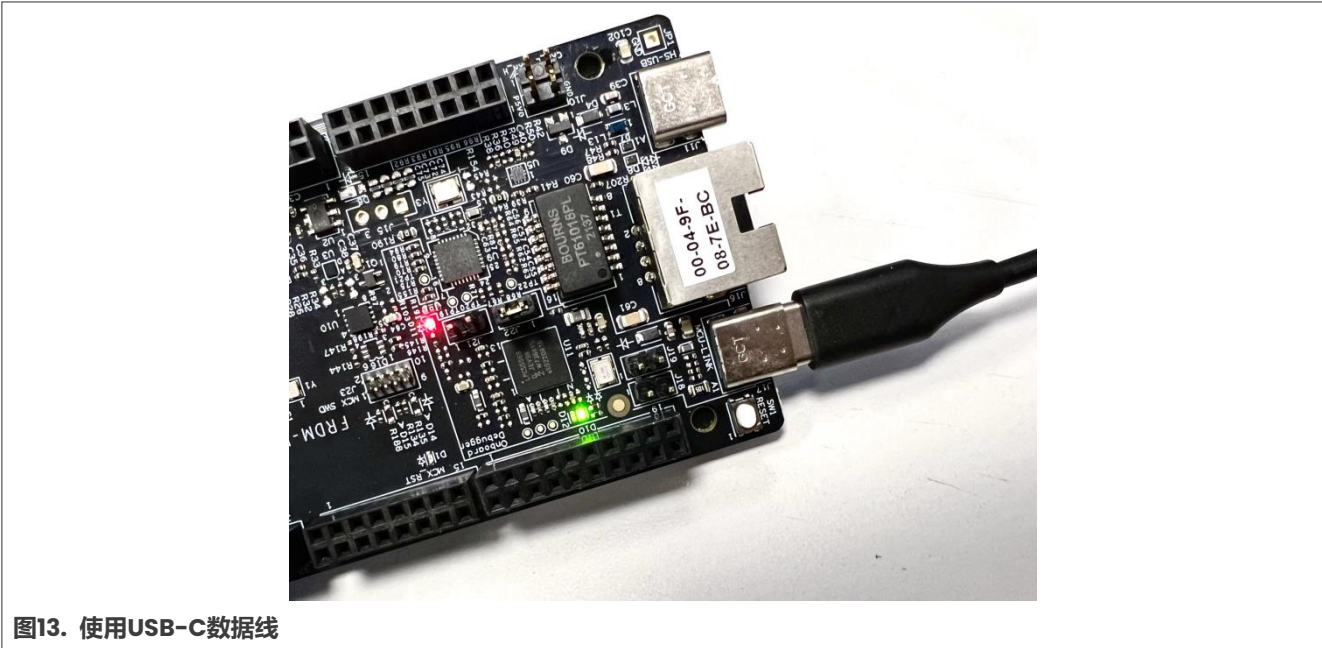


图13. 使用USB-C数据线

要安装SDK，请在[MCUxpresso SDK Builder（构建工具）](#)中选择FRDM-MCXN947，如图14所示。

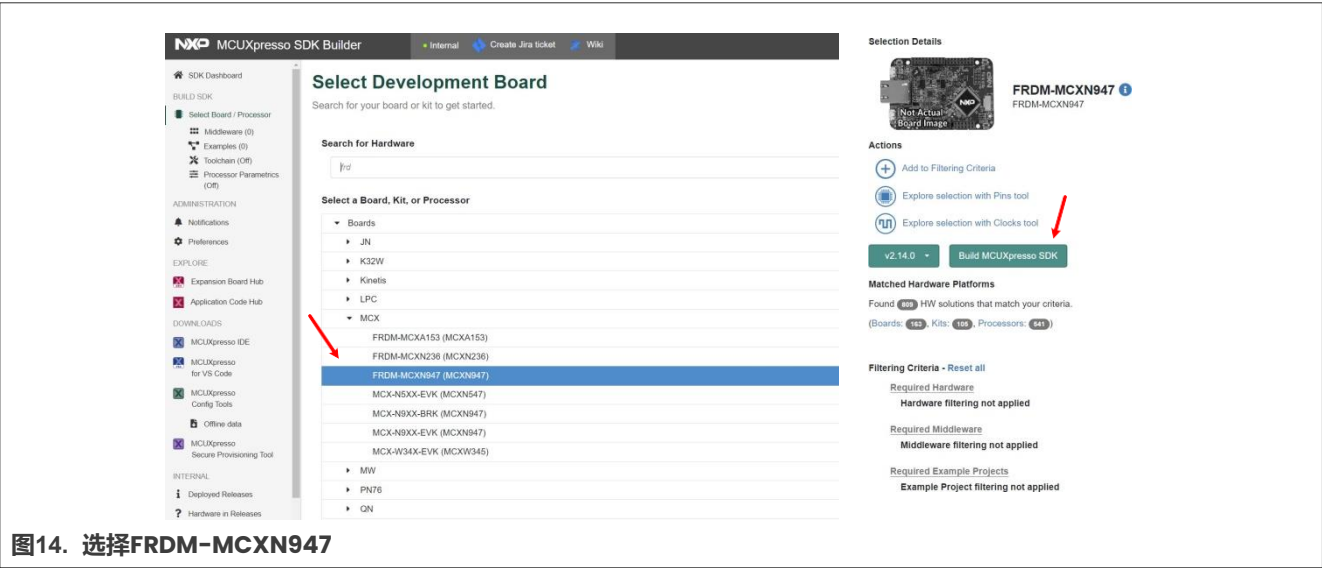


图14. 选择FRDM-MCXN947

选择eIQ中间件并点击“下载SDK”，如图15所示。

在MCX N947上利用神经处理单元（NPU）加速人脸检测演示

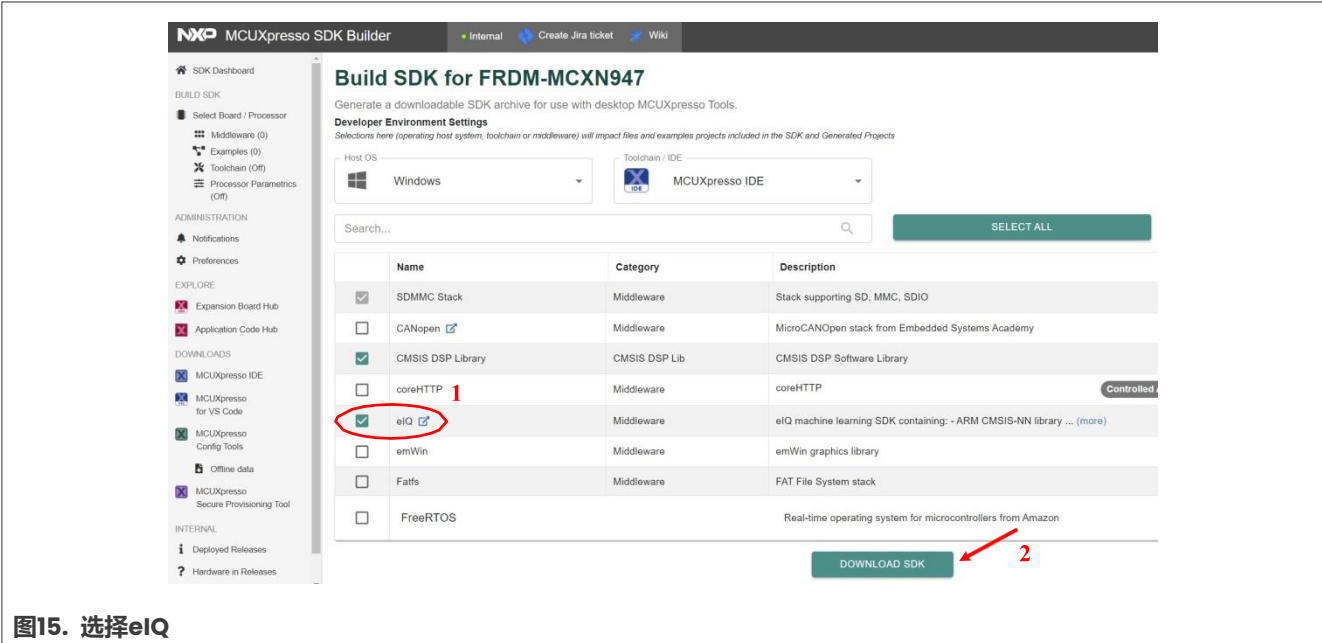


图15. 选择eIQ

完成SDK构建后，下载SDK压缩文件包，如图16所示。

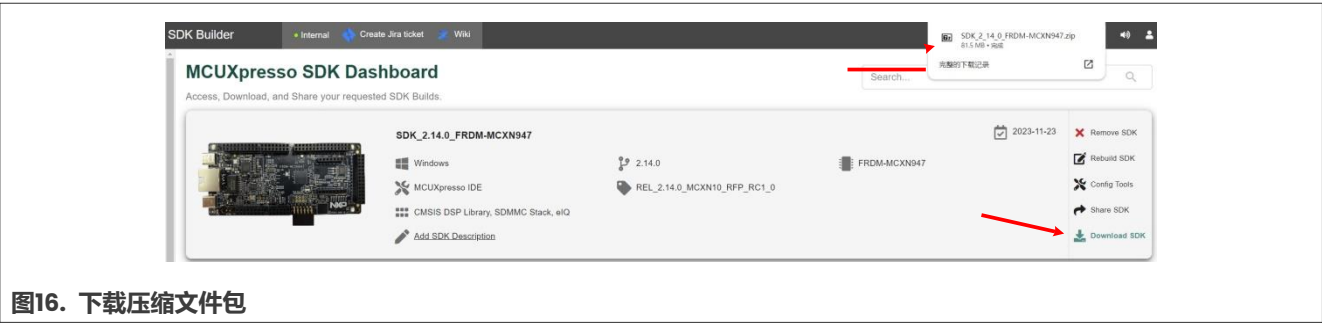


图16. 下载压缩文件包

然后打开MCUXpresso。通过将文件拖放到MCUXpresso IDE的SDK安装窗口中，可以方便地安装SDK，如图17所示。

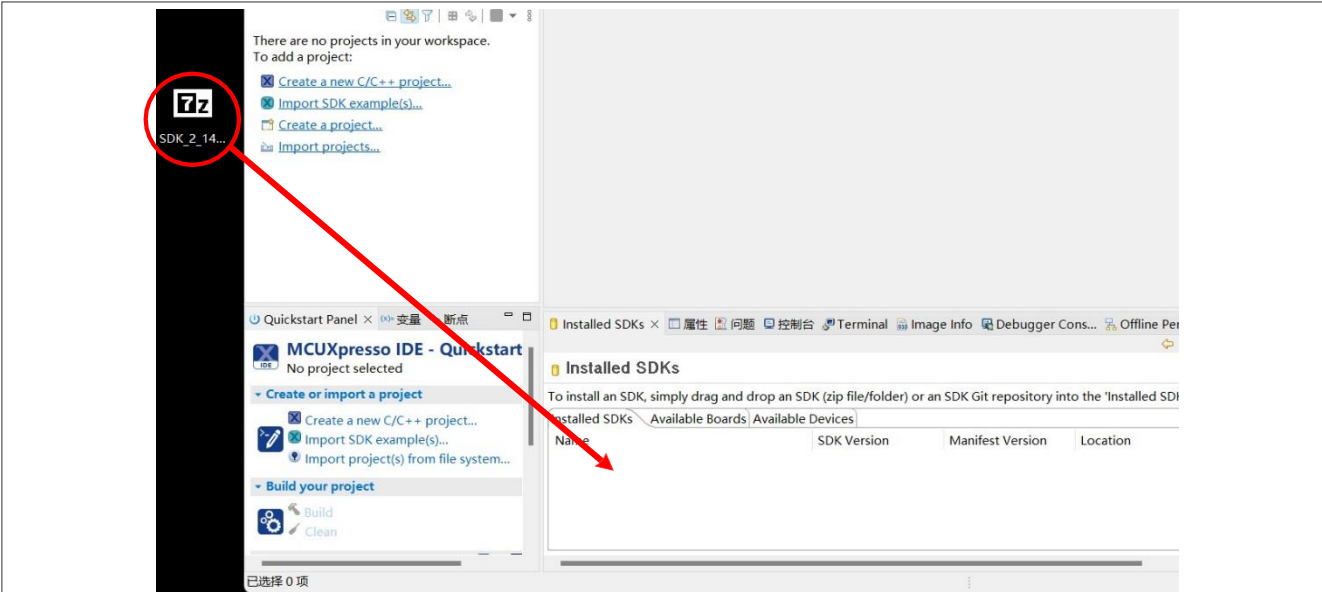


图17. 拖放文件

要导入下载的项目，请按照图18所示操作。

1. 点击**Application Code Hub**导入。
2. 搜索**multiple face detection on mcxn947**示例并将其克隆到本地工作区。

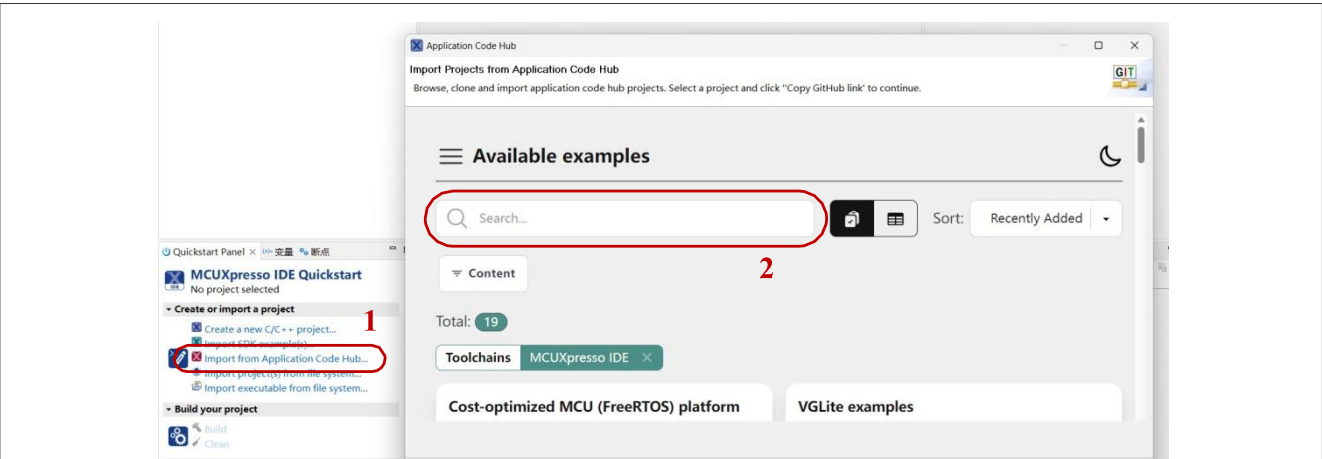


图18. 导入下载的项目

按照图19所示的步骤操作。

1. 点击**Build**。
2. 完成编译后，使用GUI Flash Tool（图19中的2）将程序写入到板子中。

在MCX N947上利用神经处理单元（NPU）加速人脸检测演示

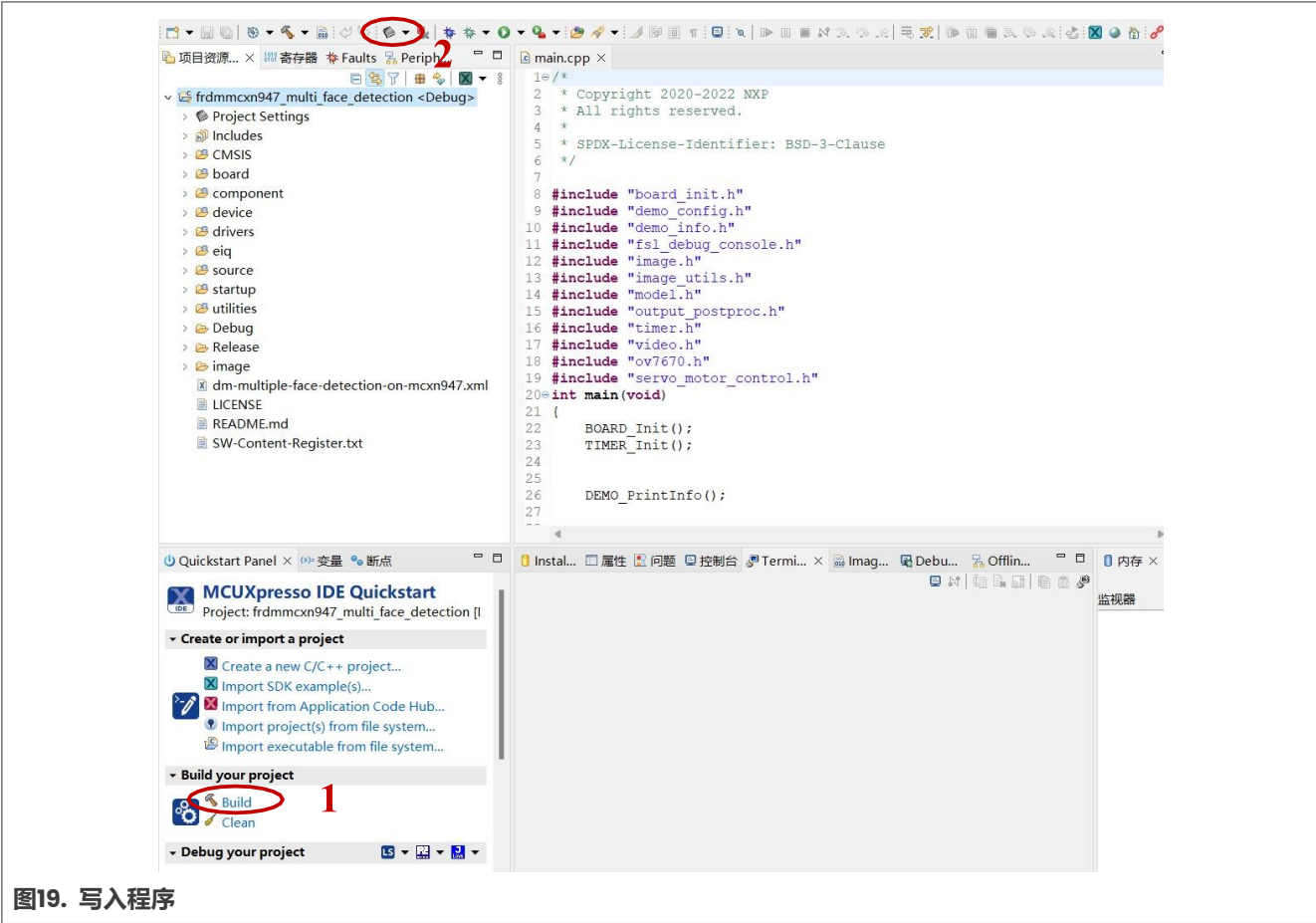


图19. 写入程序

在VS Code中，选择**MCUXpresso For VS code**插件，并在**快速启动面板**中点击**Application Code Hub**。搜索“**multiple face detection on mcxn947**”示例并将其克隆到本地工作区。等待完成后，该项目将显示在“**项目**”中。

在MCX N947上利用神经处理单元（NPU）加速人脸检测演示

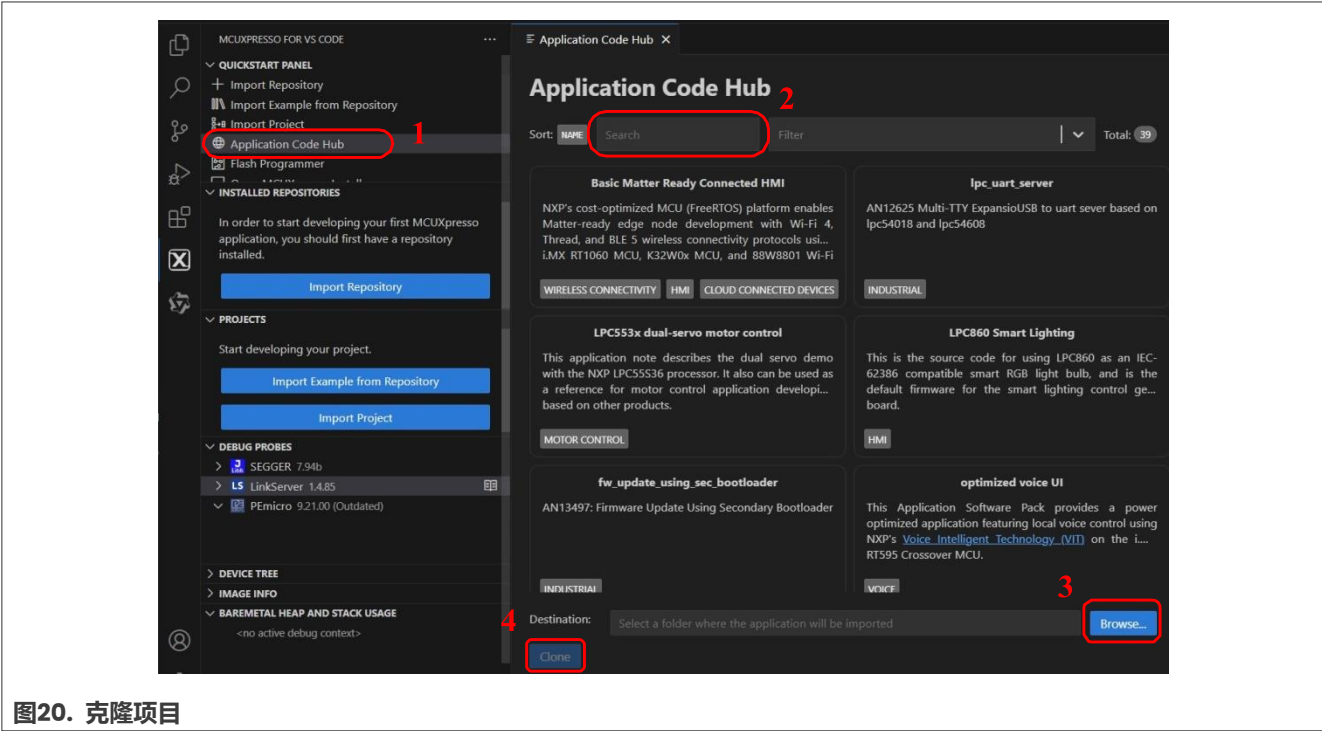


图20. 克隆项目

构建项目。完成编译后烧录板子。

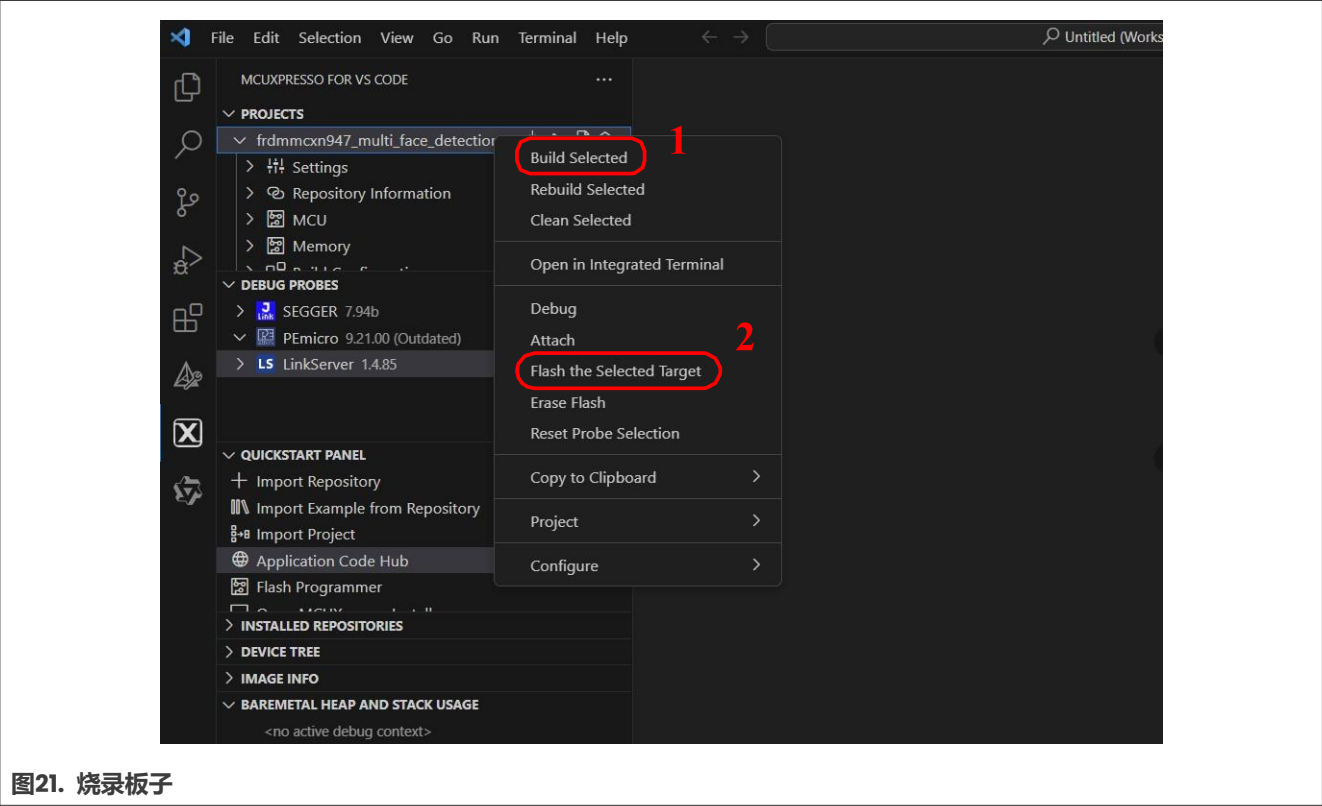


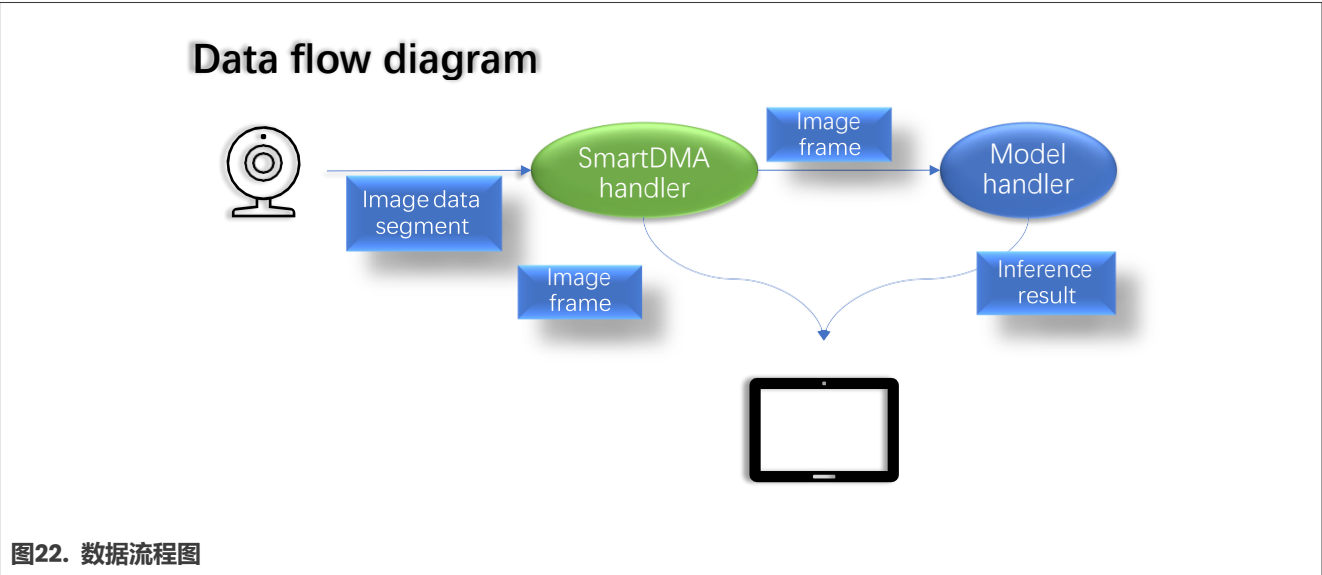
图21. 烧录板子

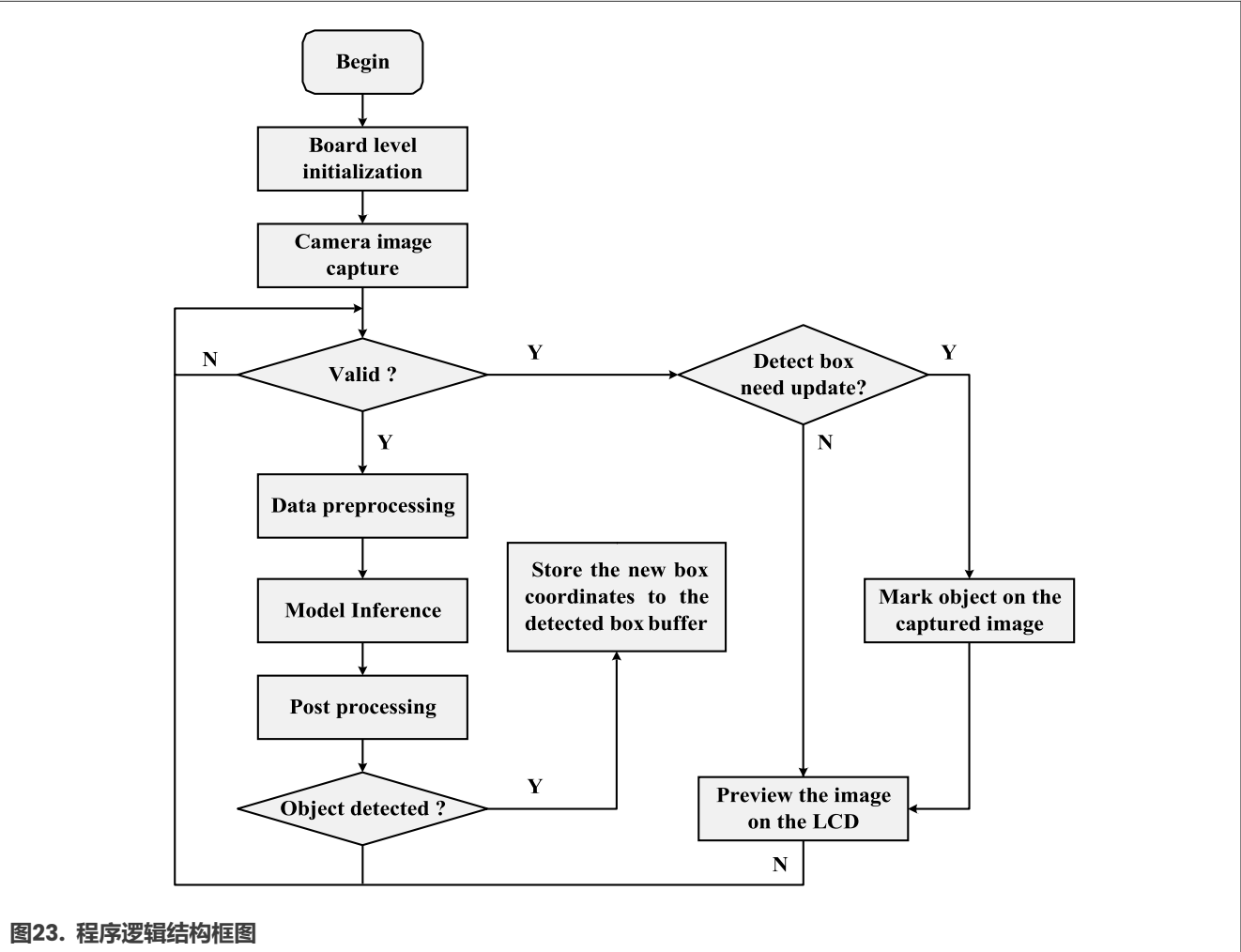
5 软件细节

5.1 程序流程

MCX N94不支持CSI接口，但我们可以改用FlexIO模拟CSI接受摄像头数据。Smart DMA（EZH）从FlexIO中读取帧数据并写入RAM。更多细节请参阅《如何在MCXN MCU中使用SmartDMA实现摄像头接口》（文档AN14191）。

图22显示了数据流程图，图23显示了程序逻辑结构框图。





器件上电后，初始化板子，包括PIN和时钟配置，接着是摄像头和LCD显示屏的初始化。然后，通过TensorFlow Lite for Microcontrollers导入模型并分配RAM资源。

系统运行时，摄像头捕获的图像经过处理后输入到模型中进行推理和后期处理。上述过程与LCD显示并行进行。当摄像头正在捕获图像时，LCD显示屏也会刷新。在LCD显示屏上显示图像之前，程序会确定是否需要在捕获的图像上叠加检测到的框（box）。如果需要，将其叠加在图像上，然后图像显示在LCD显示屏上。

5.2 模型推理和后期处理

将所捕获图像的大小调整为160×128，像素值被压缩至-1 - 1，然后被输入到模型中进行推理。如图24所示，有三种不同规模的预测输出。以左下角的第一个输出维度为例，输出维度16×20×18也可以表示为16×20×3×6。16×20表示原始图像被划分的区域数量，相当于320个方格。每个特征图对应三种尺寸的锚框（anchor box），所有网格共享这些锚框，每个锚框对应一个6维向量，即（ t_x 、 t_y 、 t_h 、 t_w 、置信度、类别），描述当前网格的推理结果。

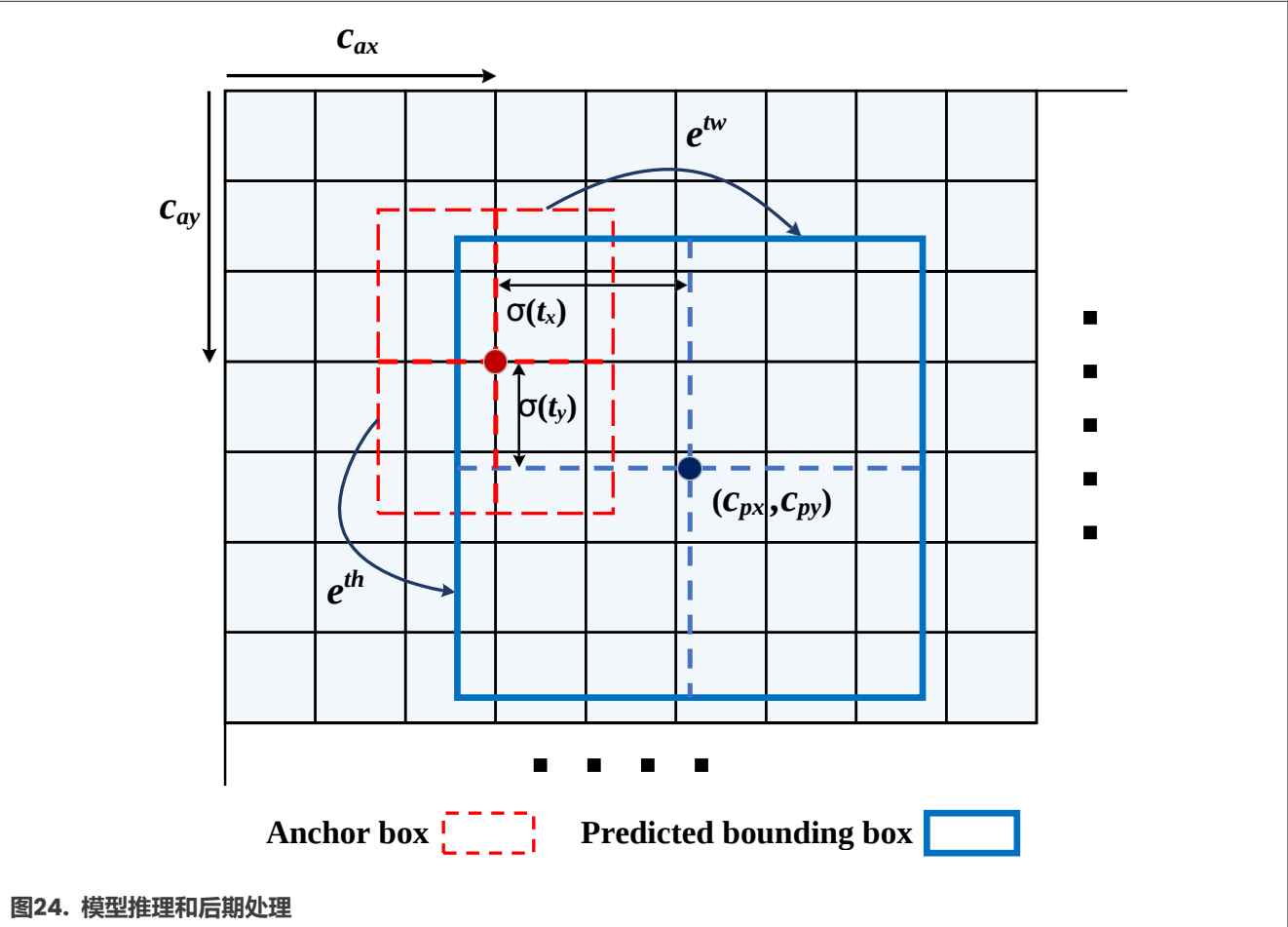


图24. 模型推理和后期处理

如图24所示，对于每个网格，基于三个锚框之一，可以根据公式(5-2)计算得出的 t_x 、 t_y 、 t_h 和 t_w ，计算预测的边界框， $\sigma(\cdot)$ 是sigmoid函数，如(5-1)所示。

(c_{ax}, c_{ay}) 是锚框中心的坐标，即对应网格的左上角。

(c_{px}, c_{py}) 是预测的边界框中心的坐标。

h_a 和 w_a 表示锚框的高度和宽度。

h_p 和 w_p 表示预测的边界框的高度和宽度。

置信度用于选择最合适的预测结果。

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

$$\begin{cases} c_{px} = c_{ax} + \sigma(t_x) \\ c_{py} = c_{ay} + \sigma(t_y) \\ h_p = h_a e^{t_h} \\ w_p = w_a e^{t_w} \end{cases} \tag{2}$$

6 结果

图25展示了在板上运行程序的结果。推理时间约为25毫秒，内存使用情况如图26所示。



图25. 运行程序的结果

Memory region	Used Size	Region Size	%age Used
PROGRAM_FLASH0:	420580 B	1 MB	40.11%
PROGRAM_FLASH1:	0 GB	1 MB	0.00%
SRAM:	405824 B	408 KB	97.14%
SRAMX:	62 KB	96 KB	64.58%
SRAMH:	0 GB	4 KB	0.00%
USB_RAM:	0 GB	4 KB	0.00%
Finished building target: frdm-mcxn947_face_tracker.axf			

图26. 内存使用

7 修订历史

表3总结了本文档的修订情况。

表3. 修订历史

文档编号	发布日期	说明
AN14132 v.1	2024年1月20日	初版发布

Legal information

Definitions

Draft — A draft status on a document indicates that the content is still under internal review and subject to formal approval, which may result in modifications or additions. NXP Semiconductors does not give any representations or warranties as to the accuracy or completeness of information included in a draft version of a document and shall have no liability for the consequences of use of such information.

Disclaimers

Limited warranty and liability — Information in this document is believed to be accurate and reliable. However, NXP Semiconductors does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information and shall have no liability for the consequences of use of such information. NXP Semiconductors takes no responsibility for the content in this document if provided by an information source outside of NXP Semiconductors.

In no event shall NXP Semiconductors be liable for any indirect, incidental, punitive, special or consequential damages (including - without limitation - lost profits, lost savings, business interruption, costs related to the removal or replacement of any products or rework charges) whether or not such damages are based on tort (including negligence), warranty, breach of contract or any other legal theory.

Notwithstanding any damages that customer might incur for any reason whatsoever, NXP Semiconductors' aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms and conditions of commercial sale of NXP Semiconductors.

Right to make changes — NXP Semiconductors reserves the right to make changes to information published in this document, including without limitation specifications and product descriptions, at any time and without notice. This document supersedes and replaces all information supplied prior to the publication hereof.

Suitability for use — NXP Semiconductors products are not designed, authorized or warranted to be suitable for use in life support, life-critical or safety-critical systems or equipment, nor in applications where failure or malfunction of an NXP Semiconductors product can reasonably be expected to result in personal injury, death or severe property or environmental damage. NXP Semiconductors and its suppliers accept no liability for inclusion and/or use of NXP Semiconductors products in such equipment or applications and therefore such inclusion and/or use is at the customer's own risk.

Applications — Applications that are described herein for any of these products are for illustrative purposes only. NXP Semiconductors makes no representation or warranty that such applications will be suitable for the specified use without further testing or modification.

Customers are responsible for the design and operation of their applications and products using NXP Semiconductors products, and NXP Semiconductors accepts no liability for any assistance with applications or customer product design. It is customer's sole responsibility to determine whether the NXP Semiconductors product is suitable and fit for the customer's applications and products planned, as well as for the planned application and use of customer's third party customer(s). Customers should provide appropriate design and operating safeguards to minimize the risks associated with their applications and products.

NXP Semiconductors does not accept any liability related to any default, damage, costs or problem which is based on any weakness or default in the customer's applications or products, or the application or use by customer's third party customer(s). Customer is responsible for doing all necessary testing for the customer's applications and products using NXP Semiconductors products in order to avoid a default of the applications and the products or of the application or use by customer's third party customer(s). NXP does not accept any liability in this respect.

Terms and conditions of commercial sale — NXP Semiconductors products are sold subject to the general terms and conditions of commercial sale, as published at <https://www.nxp.com.cn/profile/terms>, unless otherwise agreed in a valid written individual agreement. In case an individual agreement is concluded only the terms and conditions of the respective agreement shall apply. NXP Semiconductors hereby expressly objects to applying the customer's general terms and conditions with regard to the purchase of NXP Semiconductors products by customer.

Export control — This document as well as the item(s) described herein may be subject to export control regulations. Export might require a prior authorization from competent authorities.

Suitability for use in non-automotive qualified products — Unless this document expressly states that this specific NXP Semiconductors product is automotive qualified, the product is not suitable for automotive use. It is neither qualified nor tested in accordance with automotive testing or application requirements. NXP Semiconductors accepts no liability for inclusion and/or use of non-automotive qualified products in automotive equipment or applications.

In the event that customer uses the product for design-in and use in automotive applications to automotive specifications and standards, customer (a) shall use the product without NXP Semiconductors' warranty of the product for such automotive applications, use and specifications, and (b) whenever customer uses the product for automotive applications beyond NXP Semiconductors' specifications such use shall be solely at customer's own risk, and (c) customer fully indemnifies NXP Semiconductors for any liability, damages or failed product claims resulting from customer design and use of the product for automotive applications beyond NXP Semiconductors' standard warranty and NXP Semiconductors' product specifications.

Translations — A non-English (translated) version of a document, including the legal information in that document, is for reference only. The English version shall prevail in case of any discrepancy between the translated and English versions.

Security — Customer understands that all NXP products may be subject to unidentified vulnerabilities or may support established security standards or specifications with known limitations. Customer is responsible for the design and operation of its applications and products throughout their lifecycles to reduce the effect of these vulnerabilities on customer's applications and products. Customer's responsibility also extends to other open and/or proprietary technologies supported by NXP products for use in customer's applications. NXP accepts no liability for any vulnerability. Customer should regularly check security updates from NXP and follow up appropriately. Customer shall select products with security features that best meet rules, regulations, and standards of the intended application and make the ultimate design decisions regarding its products and is solely responsible for compliance with all legal, regulatory, and security related requirements concerning its products, regardless of any information or support that may be provided by NXP.

NXP has a Product Security Incident Response Team (PSIRT) (reachable at PSIRT@nxp.com) that manages the investigation, reporting, and solution release to security vulnerabilities of NXP products.

NXP B.V. — NXP B.V. is not an operating company and it does not distribute or sell products.

Trademarks

Notice: All referenced brands, product names, service names, and trademarks are the property of their respective owners.

NXP — wordmark and logo are trademarks of NXP B.V.

在MCX N947上利用神经处理单元（NPU）加速人脸检测演示

AMBA, Arm, Arm7, Arm7TDMI, Arm9, Arm11, Artisan, big.LITTLE, Cordio, CoreLink, CoreSight, Cortex, DesignStart, DynamIQ, Jazelle, Keil, Mali, Mbed, Mbed Enabled, NEON, POP, RealView, SecurCore, Socrates, Thumb, TrustZone, ULINK, ULINK2, ULINK-ME, ULINK-PLUS, ULINKpro, μ Vision, Versatile — are trademarks and/or registered trademarks of Arm Limited (or its subsidiaries or affiliates) in the US and/or elsewhere. The related technology may be protected by any or all of patents, copyrights, designs and trade secrets. All rights reserved.

eIQ — is a trademark of NXP B.V.

MCX — is a trademark of NXP B.V.

TensorFlow, the TensorFlow logo and any related marks — are trademarks of Google Inc.

目录

1 介绍..... 2

2 NPU概述..... 2

3 人脸检测模型..... 3

4 硬件和软件环境设置..... 6

5 软件细节..... 15

5.1 程序流程..... 15

5.2 模型推理和后期处理..... 16

6 结果..... 18

7 修订历史..... 19

法律声明..... 20

Please be aware that important notices concerning this document and the product(s) described herein, have been included in section 'Legal information'.

© 2024 NXP B.V. All rights reserved.

For more information, please visit: <https://www.nxp.com.cn>
Date of release: 20 January 2024
Document identifier: AN14132