NeuroPilot-Micro MT3620 Demo

Version:	0.2 (Beta Release)
Release date:	2020-07-10

MediaTek Proprietary and Confidential. © 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited.

MT3620 Demo

Demos for the TinyML application, operating on the MT3620, are presented below. Even though the MT3620 contains two high-performance Cortex-M4 cores, these demos will further showcase how NeuroPilot Micro can assist with speed and memory optimization.

The demos include the following:

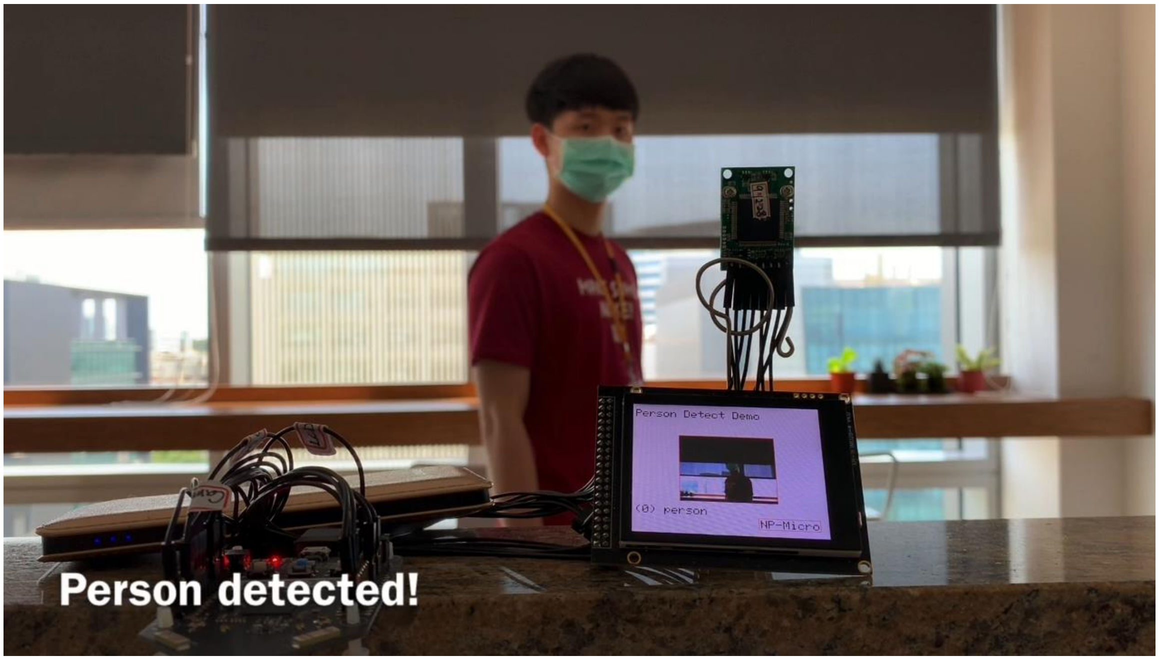
- Tiny Vision Demo - Showcases the feasibility of real-time vision detection
- Dual-Core Processing Demo - Demonstrates how to achieve performance boosts for multiple neural network models
- Handwriting Recognition Demo - How to get started, with a neural network example
- Neural Network Models - Provides a quick reference guide for neural network models

All demos can be reproduced by NeuroPilot-Micro SDK with the proposed platform.

Tiny Vision Demo

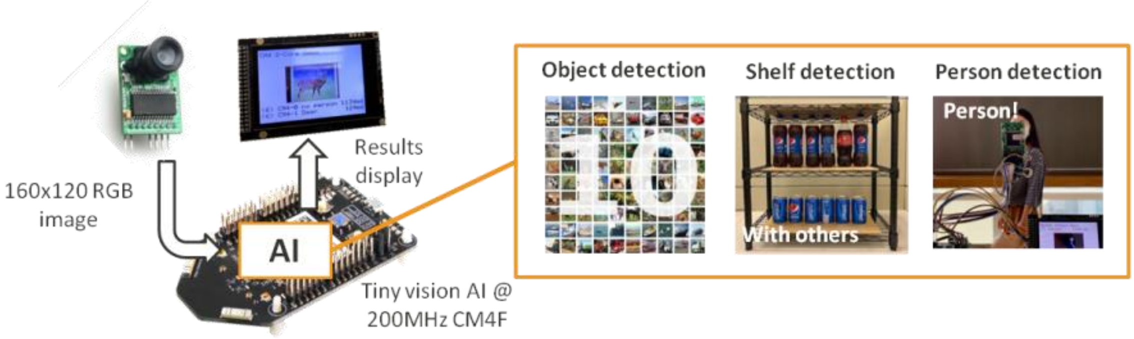
This demo will showcase the feasibility of real-time vision detection on the MT3620. Given that the Person Detection model is too large (250KB) to run with TCM (192KB max), with NeuroPilot-Micro optimization the model can use serial Flash (up to 1MB), with a **dramatically** reduced detection time of **around 1 second**, which is comparable to the performance of TCM alone.

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited



Person Detection Demo

The person detection neural network model is based on TensorFlow-Lite for Microcontrollers Example. The QVGA (160x120) image is captured by camera and converted to 96x96 gray image for person detection input.

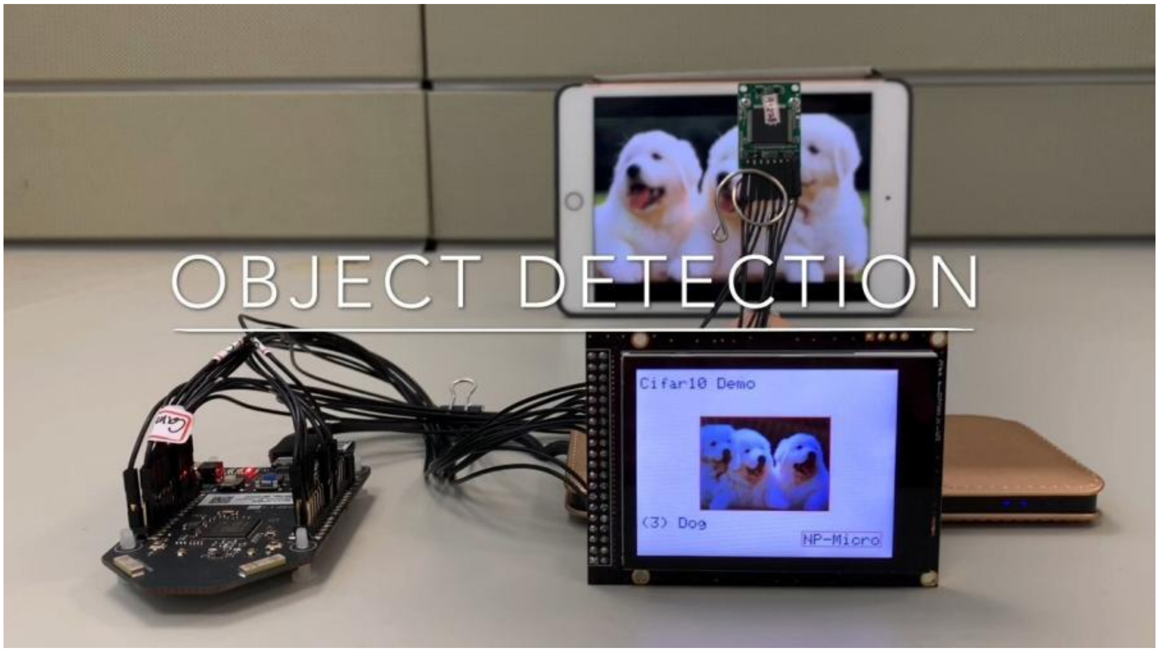


Vision-based detection offers numerous benefits for products such as mobile phones, smart cameras, and even in automotive applications, especially when interacting with edge devices. In the past, it has been quite challenging to realize the benefits of TinyML due to computing and memory limitations. With the evolution of microcontrollers, though, computing power has advanced to the point where tiny vision algorithms can now be enabled.

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited

The following tiny vision demo will showcase how to use the MT3620, in conjunction with NeuroPilot-Micro SDK, for enabling vision-related TinyML applications.

Object Detection Demo



The tiny vision demo also provides a popular object detection demo with the CIFAR-10 model, which can classify 10 objects via 32x32 RGB image. The model is trained based on TensorFlow with CIFAR-10 Dataset. To achieve higher accuracy, the model size is around 240KB. With NeuroPilot-Micro optimization, this neural network model can execute on Flash memory as fast as TCM.

Next Steps

Getting Started reference will allow you to reproduce the demo and learn more about NeuroPilot-Micro. If you already have a familiarity with neural network inference, the NeuroPilot-Micro User Guide will also serve as a helpful tool in learning more about the implementation ease.

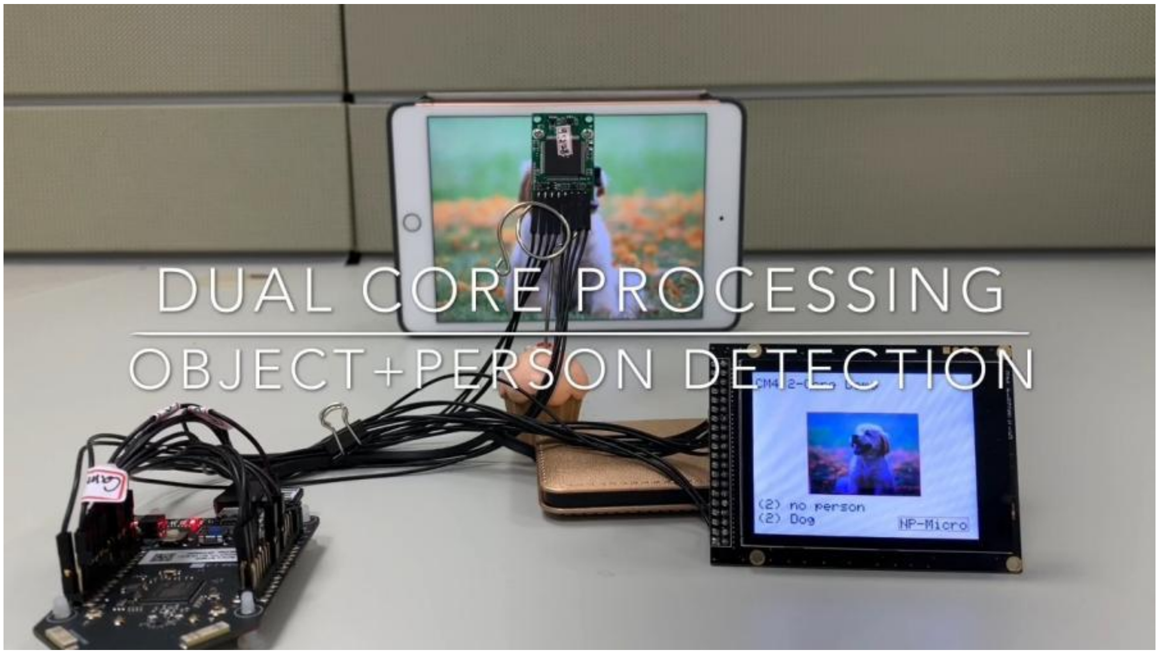
Learn how to reproduce the demo: Getting Started with Tiny Vision Demo
How to assemble the platform: Learn More: Tiny Vision Platform
Learn more about NeuroPilot-Micro: Getting Started with NeuroPilot-Micro

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited

Dual-Core Processing Demo

This Dual-Core Processing demo shows the real-time performance benefits from **two Cortex-M4** cores running concurrently. Normally, when running two models on only one core, fast object detection can take longer than 1 second due to delays induced by the person detection model. When both cores are used to run each model **simultaneously**, this limitation can be avoided.

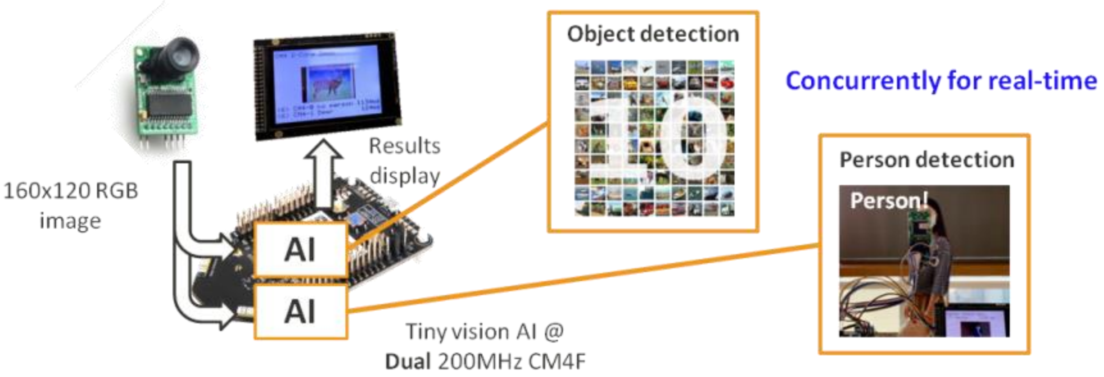
The demo below showcases the Person Detection Model and Object Detection Model running at the same time on separate Cortex-M4 cores. The result shows the fast object detection can detect objects with an almost instantaneous response time.



Another benefit worth highlighting relates to memory usage. Since each core has an exclusive 192KB TCM for running models, NeuroPilot-Micro can assist with big models by executing two models on Flash memory with performance similar to TCM alone. Thus, the flexibility of a two-core system can assist greatly in performance optimization.

It should also be noted that since the dual-core demo platform is based on the tiny vision demo platform, there are additional benefits to be gained. Specifically, with the example of the two models running on concurrent dual cores, and with the same QVGA (160x120) camera input, it is now possible to have two algorithms perform detection at the same time and showcase the results separately.

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited



Next Steps

Getting Started reference will allow you to reproduce the demo and learn more about NeuroPilot-Micro. If you already have a familiarity with neural network inference, the NeuroPilot-Micro User Guide will also serve as a helpful tool in learning more about the implementation ease.

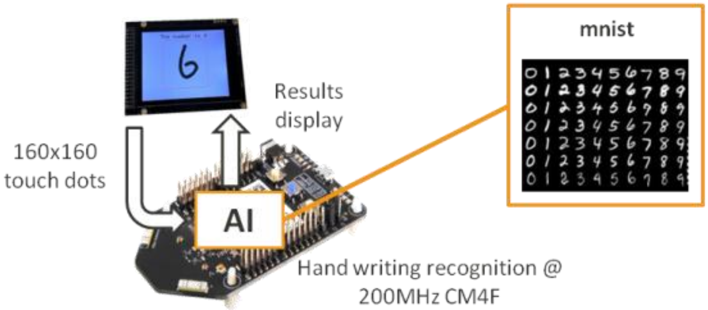
Learn how to reproduce dual-core processing demo: Getting Started with Dual-Core Processing Demo
How to assemble the platform: Learn More: Tiny Vision Platform
Learn more about NeuroPilot-Micro: Getting Started with NeuroPilot-Micro

Handwriting Recognition Demo

Handwriting recognition of numbers (ex: 0~9) is a representative example for a neural network machine learning model. A 28x28 touch-dot array is used for number recognition and users can write numbers on the panel with real-time writing trace display for a better user experience.

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited

The handwriting demo platform includes a touch panel and MT3620 EVK. The model is trained based on TensorFlow with MNIST Dataset and executed via NeuroPilot-Micro. This demo shows not only how to execute a neural network model, but also how to develop pre-processing and user interface (touch and display). NeuroPilot-Micro also helps to execute the neural network model more easily by supporting TensorFlow Lite for a microcontroller inference framework.



Next Steps

Getting Started reference will allow you to reproduce the demo and learn more about NeuroPilot-Micro. If you already have a familiarity with neural network inference, the NeuroPilot-Micro User Guide will also serve as a helpful tool in learning more about the implementation ease.

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or
in part, is strictly prohibited

Learn how to reproduce handwriting recognition demo: Getting Started with Handwriting Recognition Demo
How to assemble the platform: Learn More: Handwriting Platform
Learn more about NeuroPilot-Micro: Getting Started with NeuroPilot-Micro

Neural Network Models

Pre-built models are provided for quick reference. **".tflite"** file can be download directly for inference. The corresponding training web-site link is also provided. For easy development, we also list model size, min. tensor arena size and input/output shape. Developer can have a quick reference for deploying model(s).

There are suggestions for optimizing memory usage:

- Model is suggested to place on the external memory (ex: Flash memory) with NeuroPilot-Micro optimization.
- Tensor arena should be shared or reused.

Model name	Model size	Min. tensor arena size	Shape (In/Out)	Description
MNIST	41 KB	14x1024	28x28x1/10	Image classification of handwritten digits.(Training)
Object detection (CIFAR-10)	229 KB	45x1024	32x32x3/10	Image classification of ten objects.(Training)
Person detection	231 KB	71x1024	96x96x1/3	Classify if a person in the image or not.(Training)
Micro speech	231 KB	20x1024	49x40x1/4	Recognize 2 keywords, "yes" and "no", from speech data.(Training)

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited

Model name	Model size	Min. tensor arena size	Shape (In/Out)	Description
Magic wand	19 KB	60x1024	128x3x1/4	Detect the gesture by the linear accelerator.(Training)
Hello world	2.5 KB	676	1/1	Simulate a sine wave.(Training)

A model graph is provided as well as a helpful and **easy-to-use** online tool, called Netron.

Next Steps

For knowledge on how to execute these models, please refer to Getting Started for direction. If you already have a familiarity with neural network inference, the NeuroPilot-Micro User Guide will also serve as a helpful tool to learn more about the implementation ease.

Learn more about NeuroPilot-Micro: Getting Started with NeuroPilot-Micro

© 2020 MediaTek Inc. All rights reserved.
Unauthorized reproduction or disclosure of this document, in whole or in part, is strictly prohibited