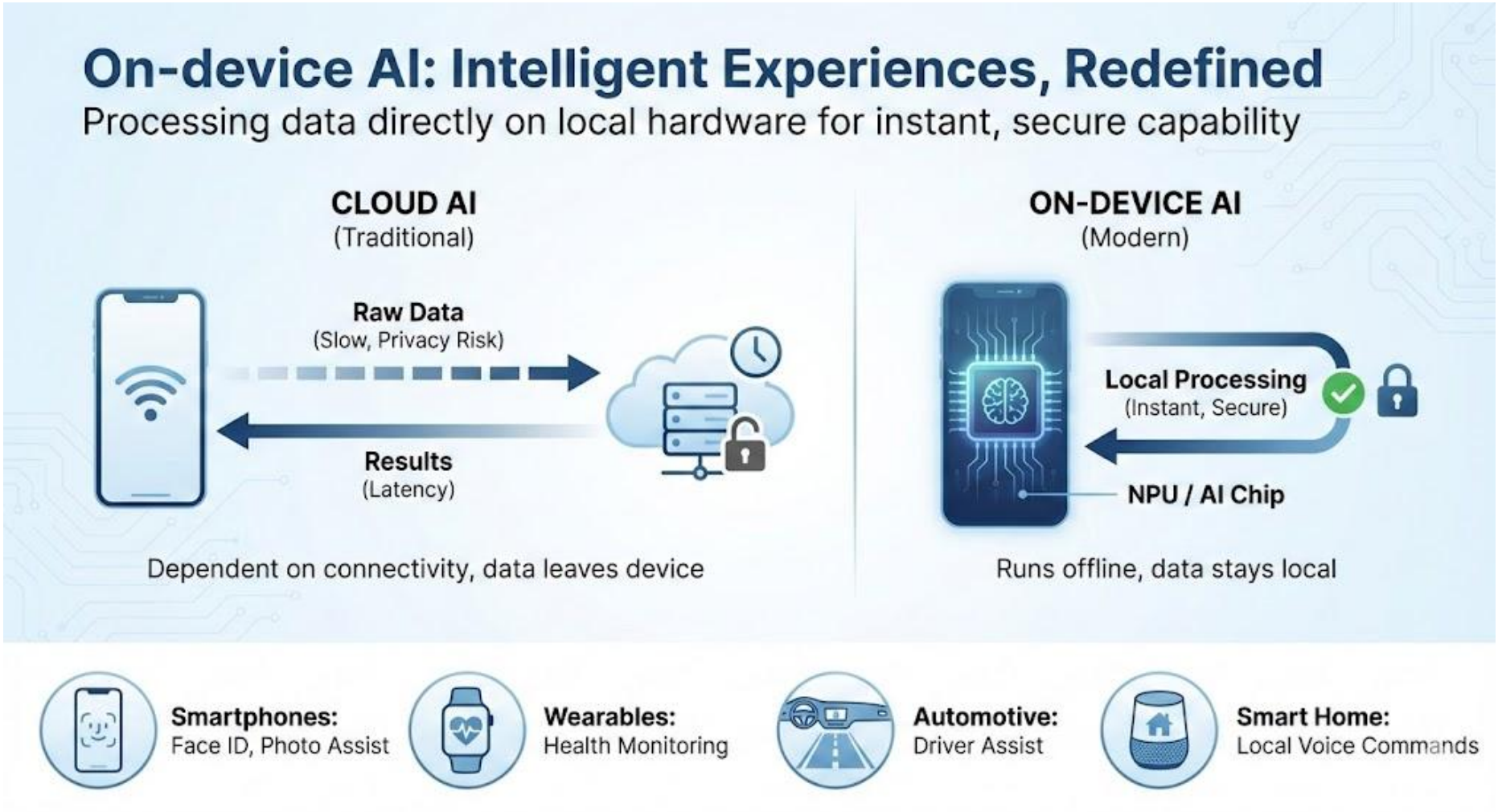
Tutorial: On-device AI

CS4262/5462: Machine Learning Systems

Junyi Shen
junyis@u.nus.edu

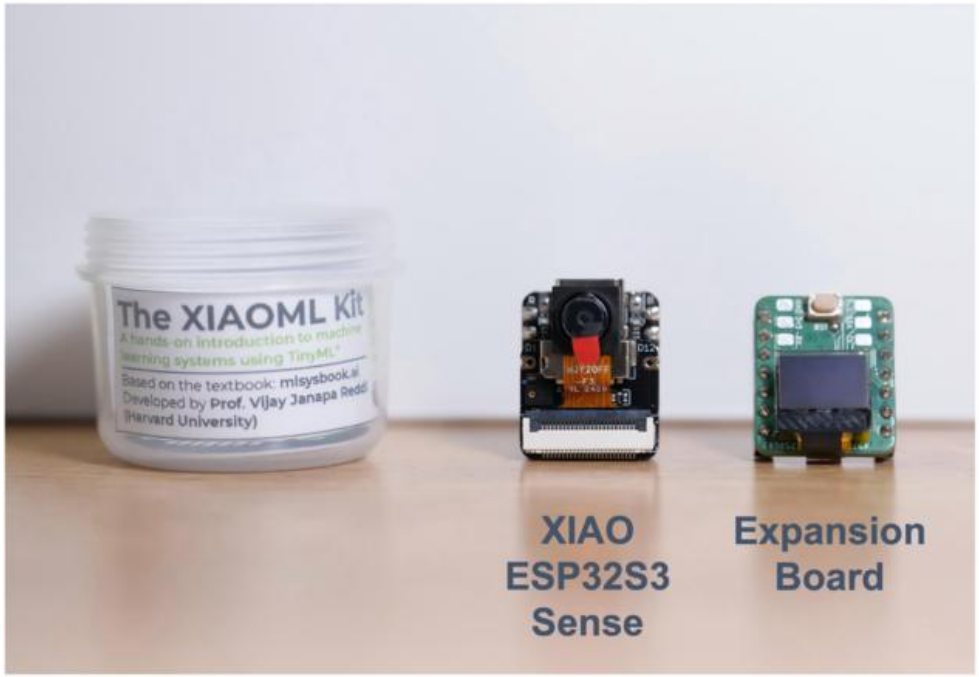
17 Feb 2026

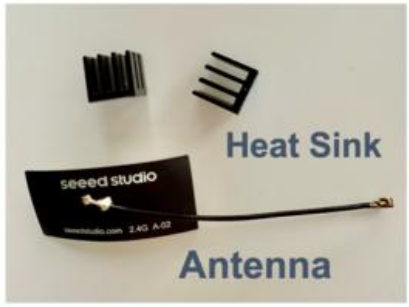
01 On-device AI



01 Device

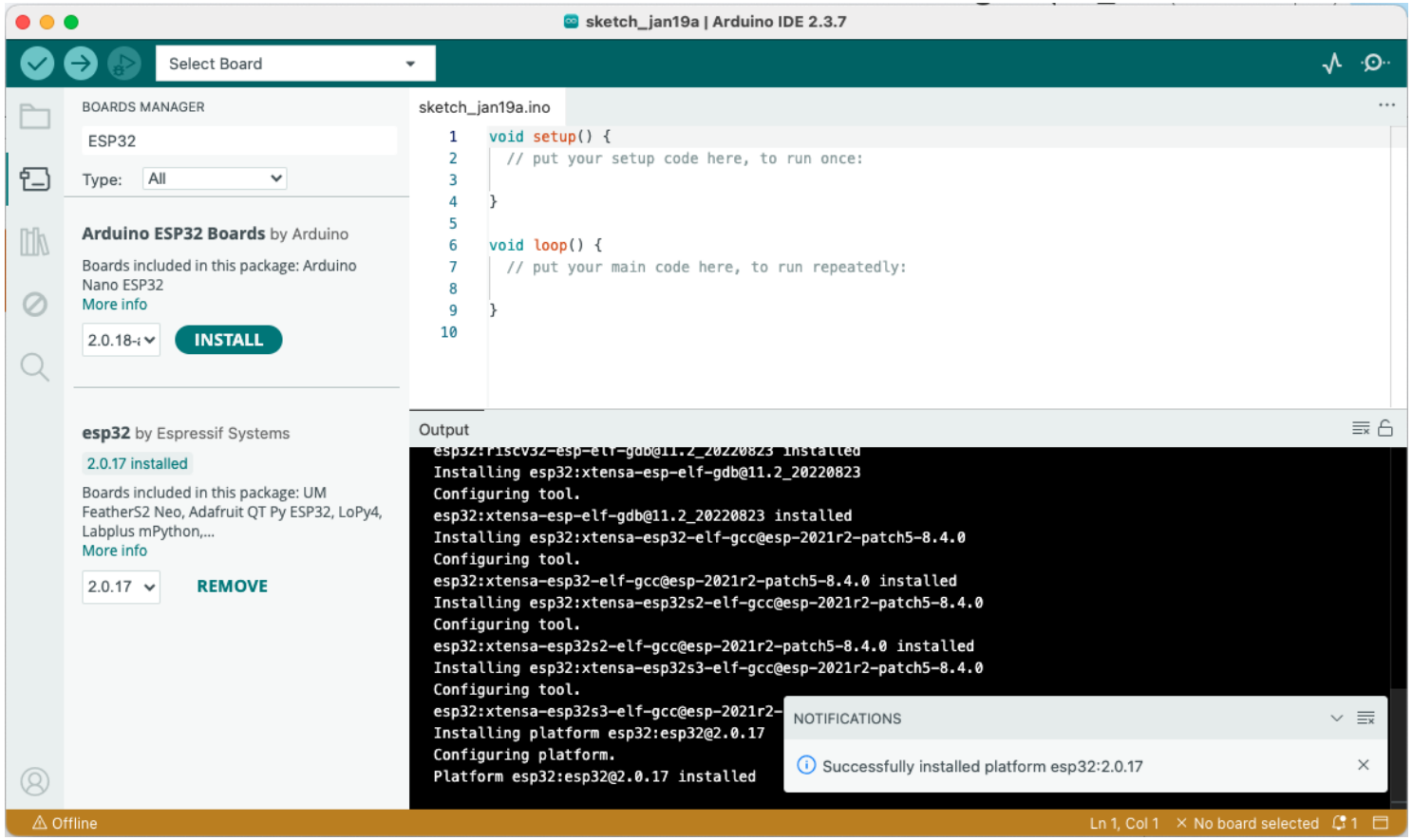
- **XIAO ESP32S3 Sense:** Main development board with integrated camera sensor, digital microphone, and SD card support
- Expansion Board: Features a 6-axis IMU (LSM6DS3TR-C) and 0.42” OLED display for motion sensing and data visualization
- SD Card Toolkit: Includes SD card and USB adapter for data storage and model deployment
- USB-C Cable: For connecting the board to your computer
- Antenna and Heat Sinks



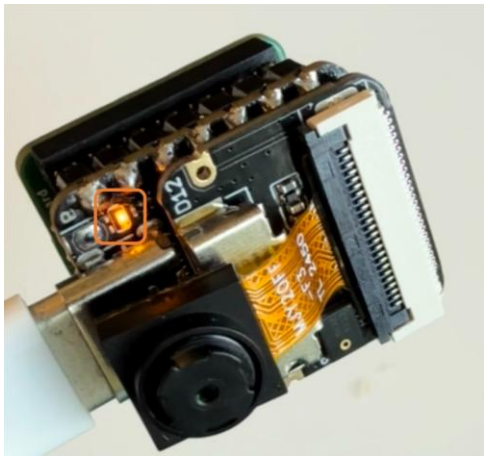


01 Install

We use Arduino for edge-AI programming.



01 BLINK



```
#define LED_BUILT_IN 21 // This line is optional

void setup() {
  pinMode(LED_BUILT_IN, OUTPUT); // Set the pin as output
}
```

Run at Start

```
// Remember that the pins work with inverted logic
// LOW to turn on and HIGH to turn off
void loop() {
  digitalWrite(LED_BUILT_IN, LOW); //Turn on
  delay (1000); //Wait 1 sec
  digitalWrite(LED_BUILT_IN, HIGH); //Turn off
  delay (1000); //Wait 1 sec
}
```

Run in Loop

Note that the pins operate with inverted logic: LOW turns on and HIGH turns off.

01 Microphone

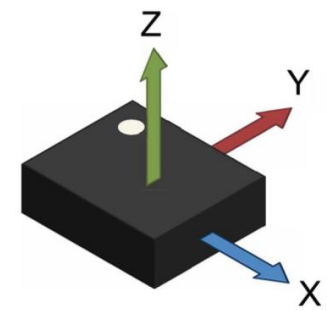
XIAOML Toolkit equipped with a digital microphone.

https://github.com/Mjrovai/XIAO-ESP32S3-Sense/tree/main/Mic_Test/XiaoEsp32s3_Mic_Test

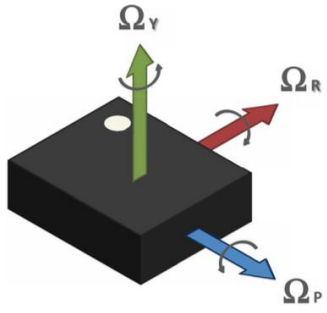
01 IMU Sensor

An Inertial Measurement Unit (IMU) is a sensor that measures motion and orientation. The LSM6DS3TR-C on your XIAOML kit is a 6-axis IMU, meaning it combines:

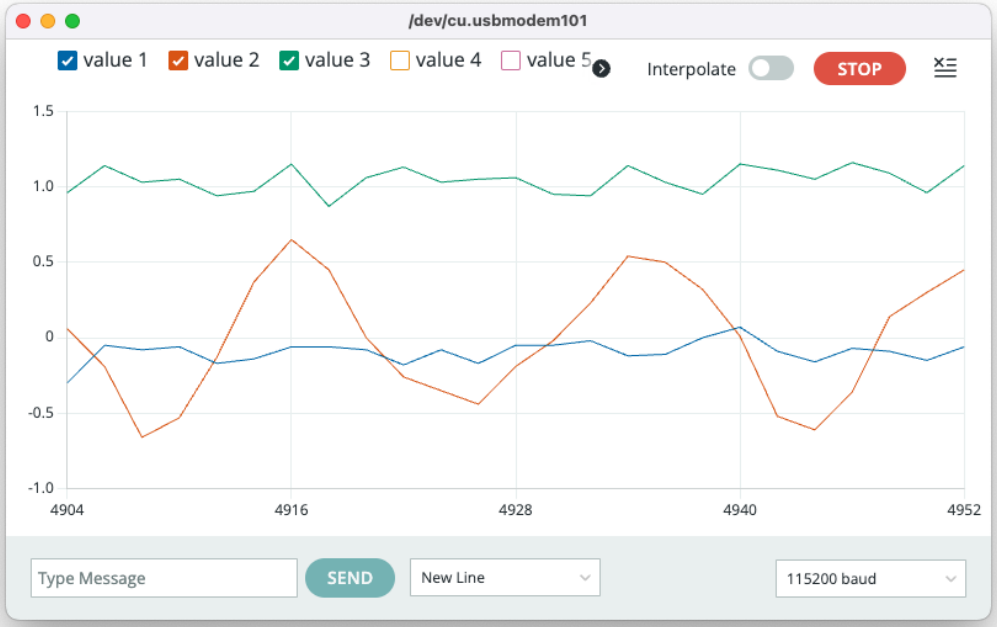
- **3-axis Accelerometer**: Measures linear acceleration (including gravity) along X, Y, and Z axes
- **3-axis Gyroscope**: Measures angular velocity (rotation rate) around X, Y, and Z axes



Direction of detectable acceleration (top view)



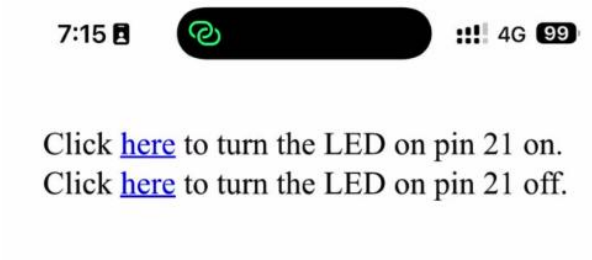
Direction of detectable angular rate (top view)



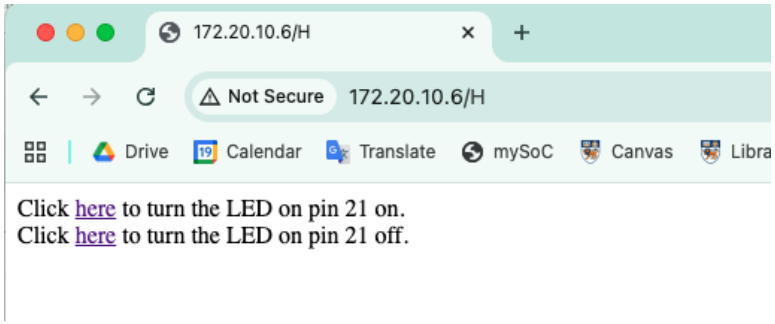
01 WIFI

WIFI Scan and remote control

```
Scan start
Scan done
36 networks found
Nr | SSID                             | RSSI | CH | Encryption
 1 | Junyi-iPhone                     |  -30 |  3 | WPA2
 2 | eduroam                          |  -56 |  3 | WPA2-EAP
 3 | NUS                              |  -57 |  3 | WPA2-EAP
 4 | NUS_STU                          |  -57 |  3 | WPA2-EAP
 5 | SOCLAB                           |  -57 |  3 | WPA+WPA2
 6 | NUS_Guest                        |  -57 |  3 | open
 7 | eduroam                          |  -59 |  9 | WPA2-EAP
 8 | NUS_STU                          |  -59 |  9 | WPA2-EAP
 9 | SOCLAB                           |  -59 |  9 | WPA+WPA2
```

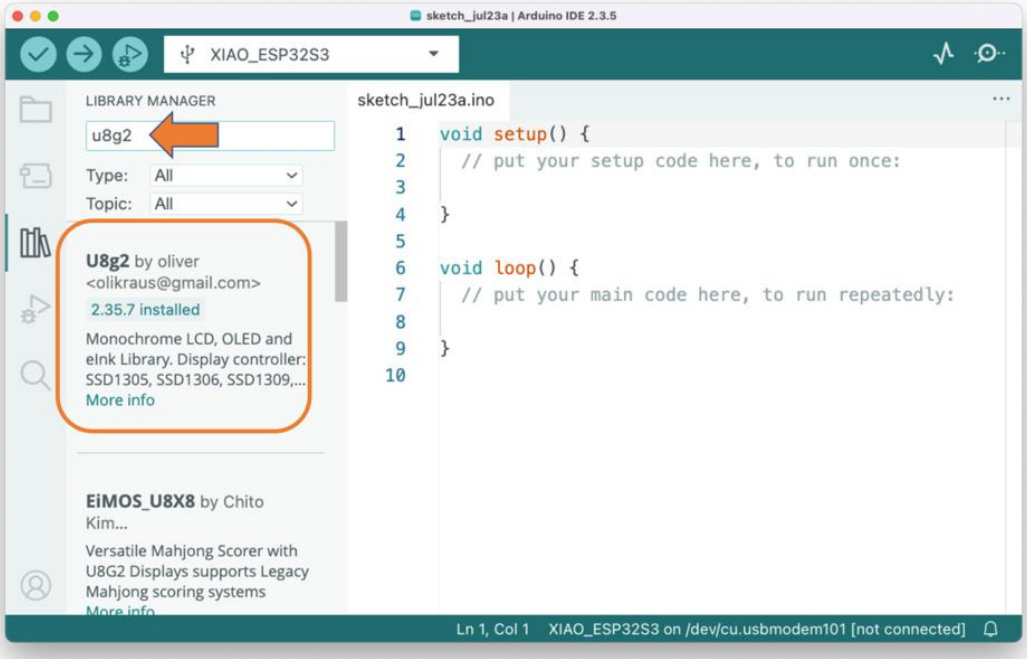


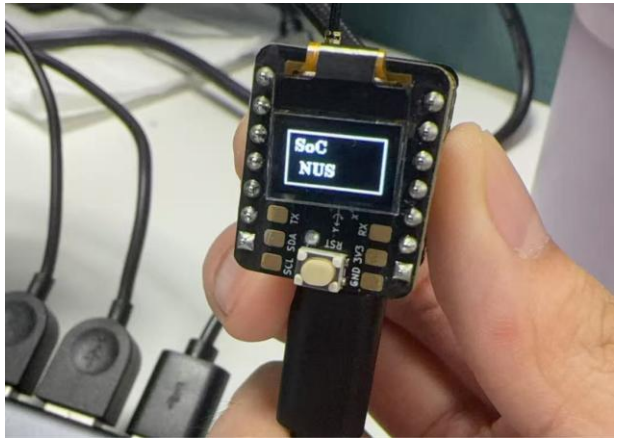

```
Connecting to Junyi-iPhone
..
WiFi connected.
IP address:
172.20.10.6
```



01 OLED

A good way to demonstrate your output – OLED





Test Code:
https://mlsysbook.ai/kits/contents/seeed/xiao_esp32s3/setup/setup.html#sec-setup-test-code-998c

02 Project Topics

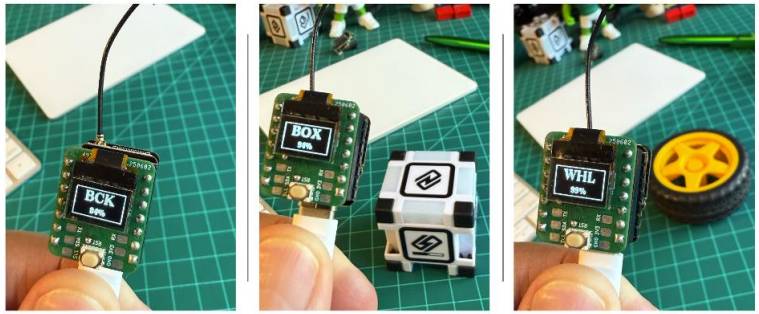
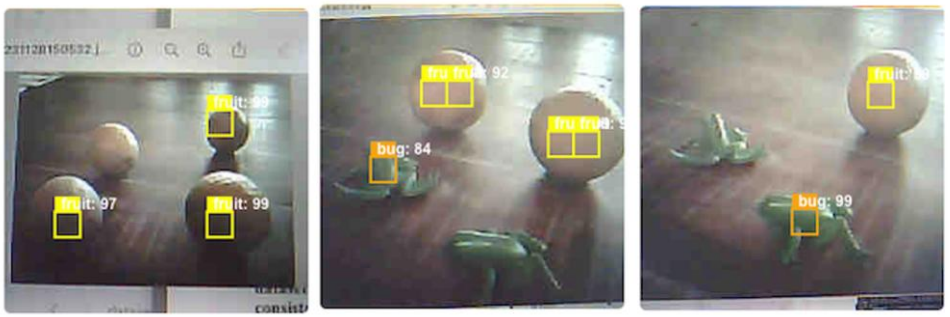
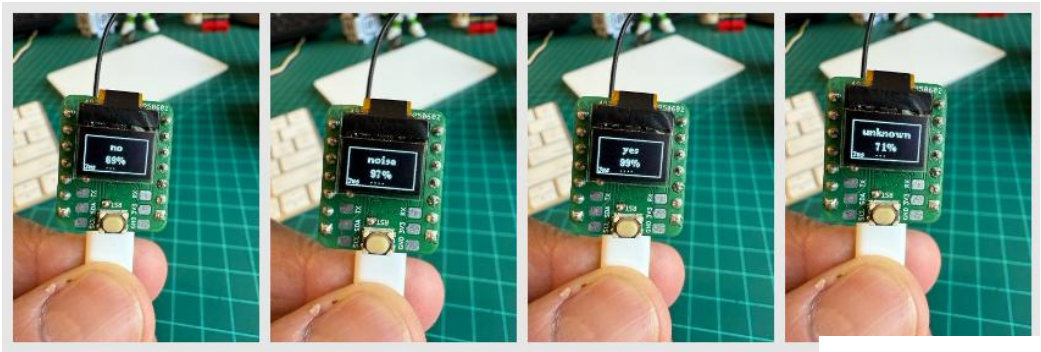


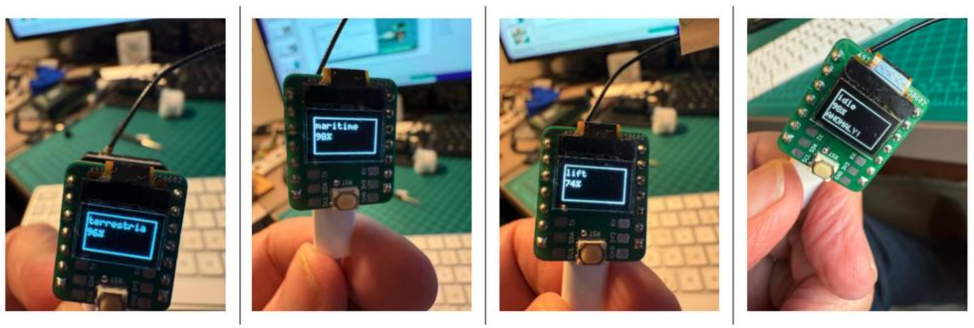
Image Classification



Object Detection



Keyword Spotting



Motion Classification
Anomaly Detection

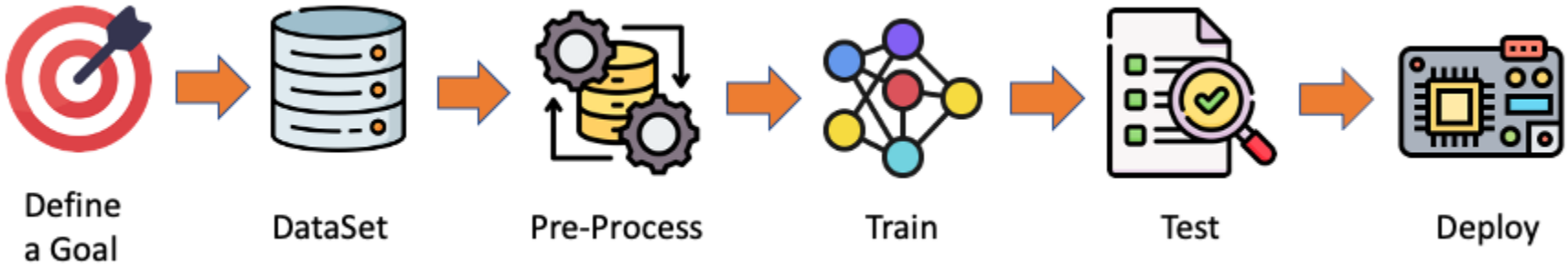
02 Project Requirements

Modality	Task	Description
Vision	Image Classification	Learn to classify images
Vision	Object Detection	Implement object detection
Sound	Keyword Spotting	Explore voice recognition systems
IMU	Motion Classification and Anomaly Detection	Classify motion data and detect anomalies

Requirements:
2 of the tutorial tasks from the above (different setting);
1 in-depth exploration with advanced post-processing

Examples:
https://mlsysbook.ai/kits/contents/seeed/xiao_esp32s3/xiao_esp32s3.html

02 Project Workflows



Arduino: On-device programming
SenseCraft AI: Easy to use, but less flexibility
Edge Impulse Studio: More powerful, seamless with Arduino

02 Project Workflows

Define
a Goal

Imagine you are creators of a startup of On-device AI…

Which setting should we focus on?

Industrial manufacturing, elderly/child caring…

What functions should we design?

Image classification, voice recognition…

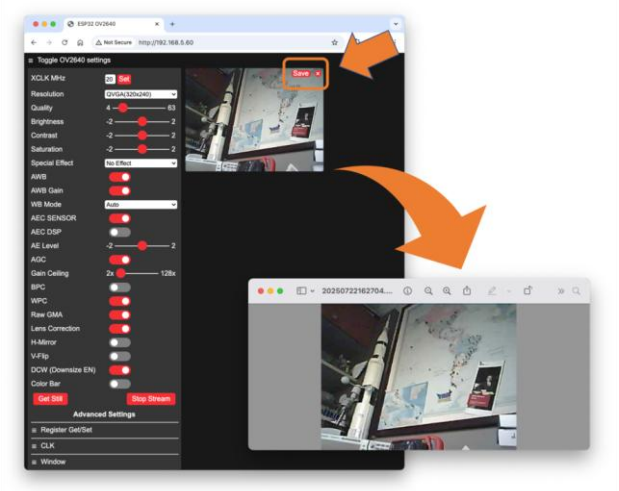
How can we implement these functions leveraging the device we have?

Camera, IMU unit, Microphone, WiFi, LED…

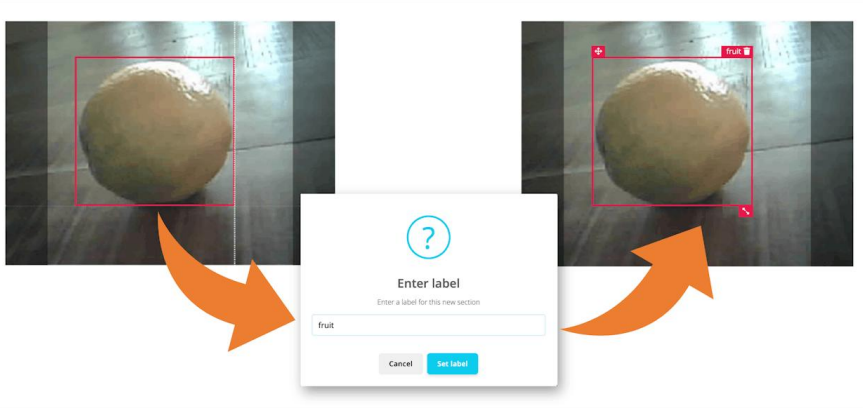
02 Project Workflows

DataSet

Gather data for your model training.



Gathering (Device/Phone)



Labeling and Uploading

https://mlsysbook.ai/kits/contents/seeed/xiao_esp32s3/object_detection/object_detection.html#sec-object-detection-collecting-dataset-xiao-esp32s3-9814

02 Project Workflows

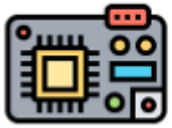
Model Choose
Hyper-parameter Config

Test your trained model

Training (On platform)

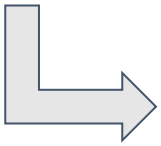
https://mlsysbook.ai/kits/contents/seeed/xiao_esp32s3/image_classification/image_classification.html#sec-image-classification-preprocessing-feature-generation-ee2e

02 Project Workflows

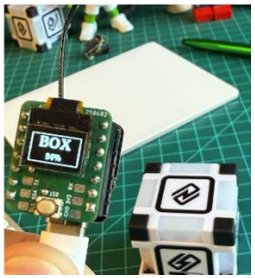


Deploy

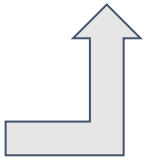
Download the trained model



Migrate to Edge device via Arduino



Test with your edge device



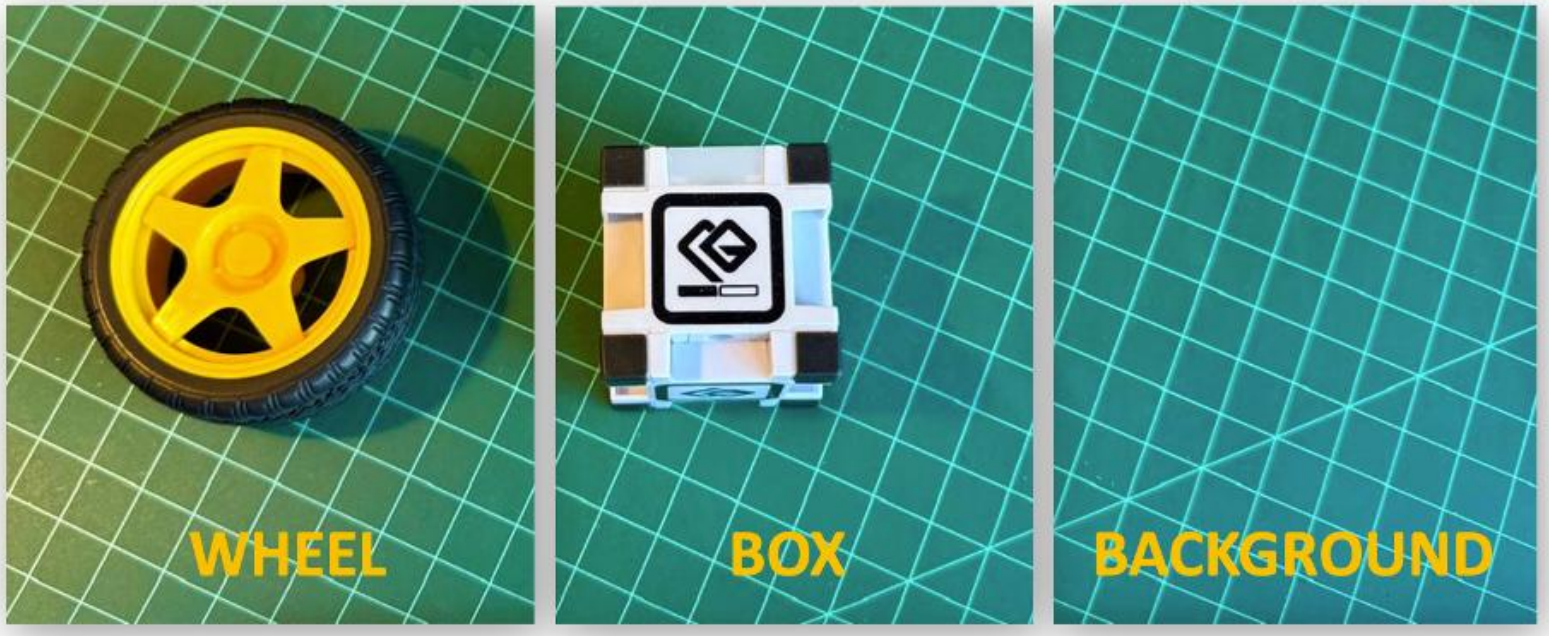
https://mlsysbook.ai/kits/contents/seeed/xiao_esp32s3/image_classification/image_classification.html#sec-image-classification-model-deployment-arduino-library-ei-studio-5579

02 Project Example

An industrial installation that should automatically sort wheels and boxes.

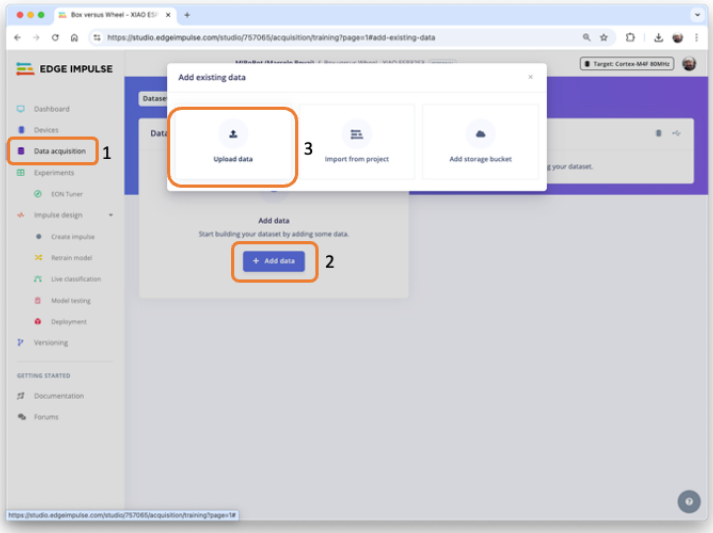
02 Project Example

An industrial installation that should automatically sort wheels and boxes.

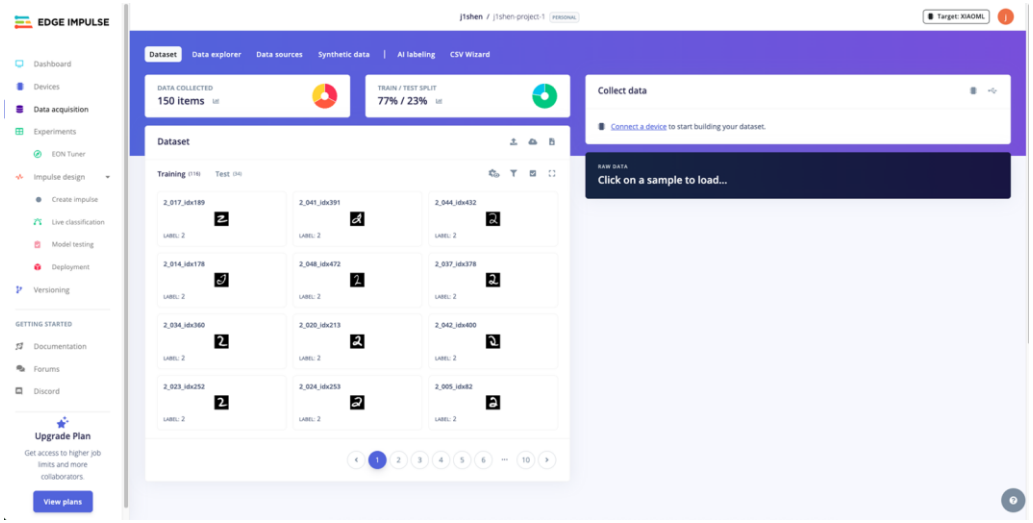


02 Project Example

Upload your dataset to the EDGE IMPULSE Platform

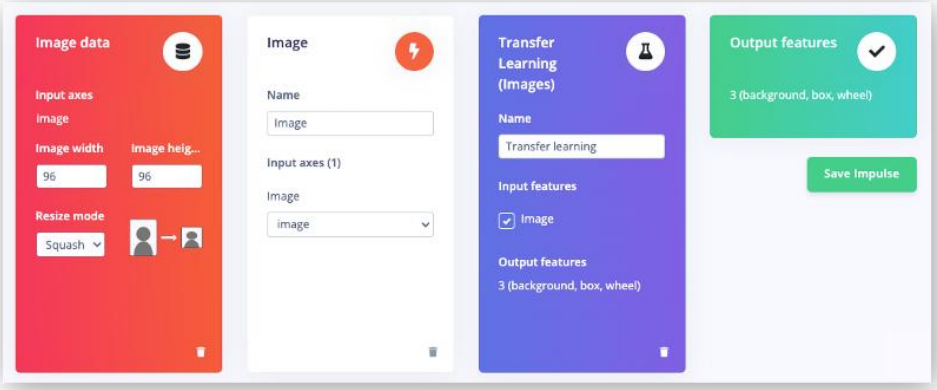


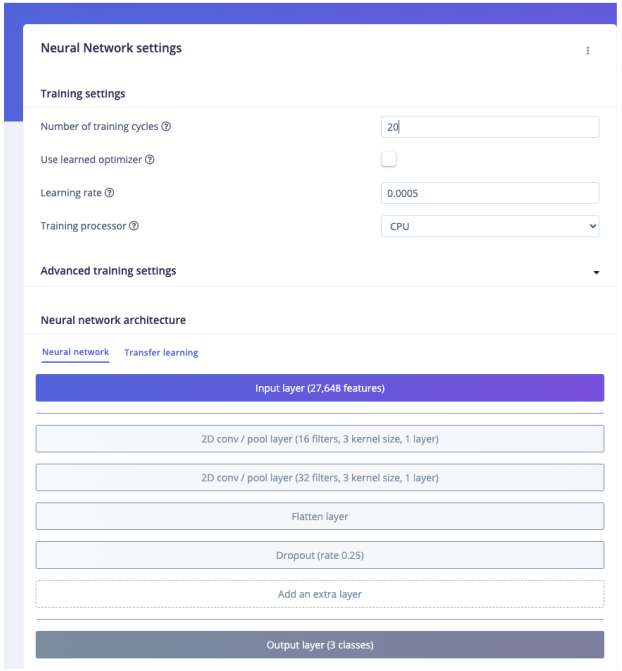
Labeling



02 Project Example

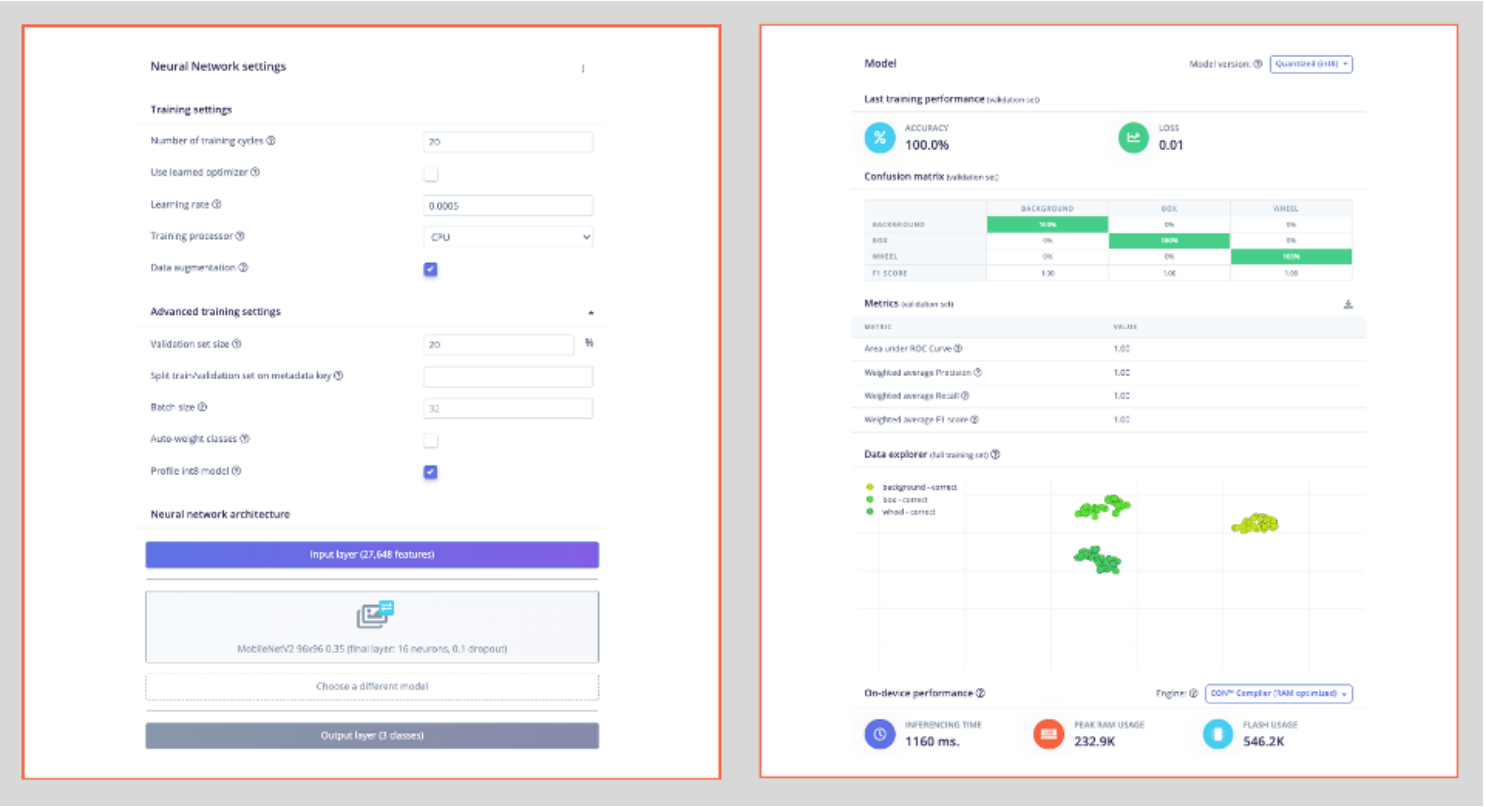
Define training model and parameters





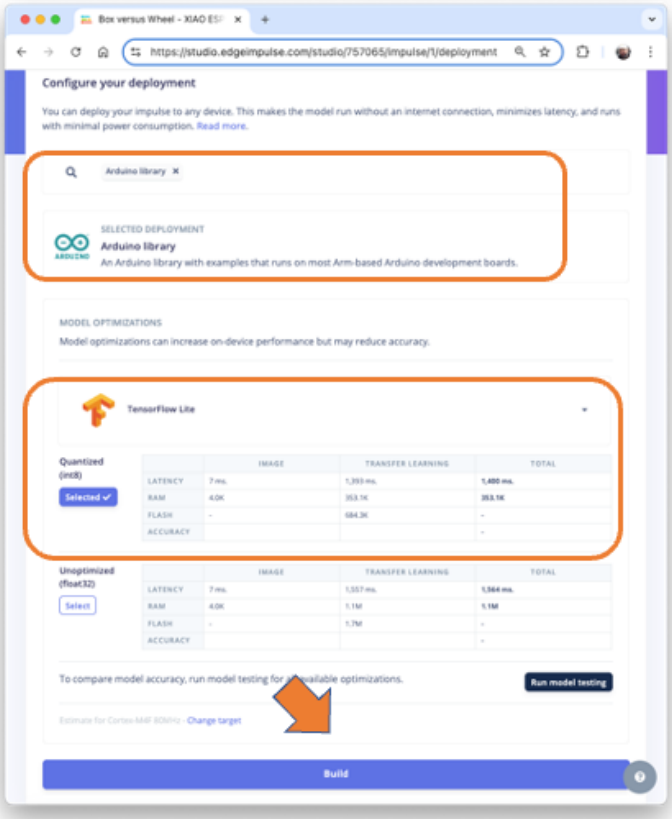
02 Project Example

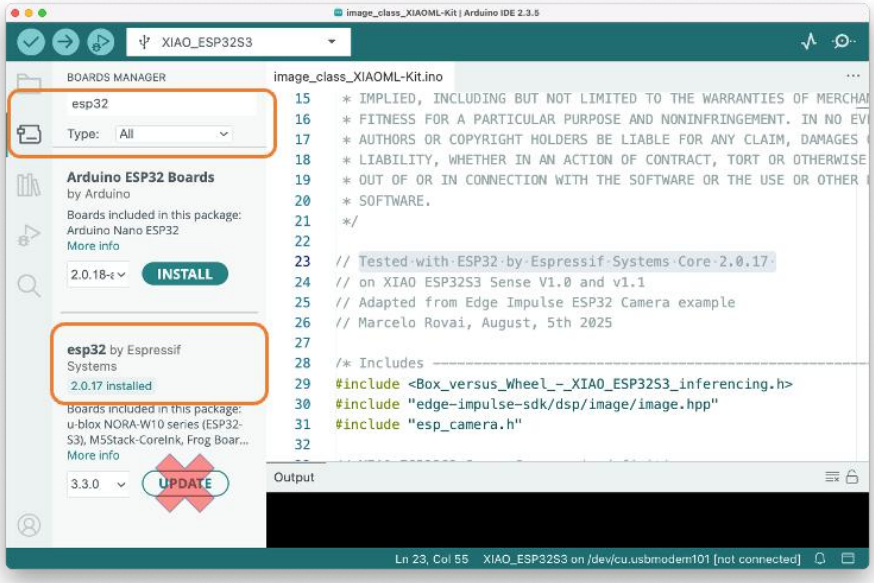
Train/Test the model and get the report.



02 Project Example

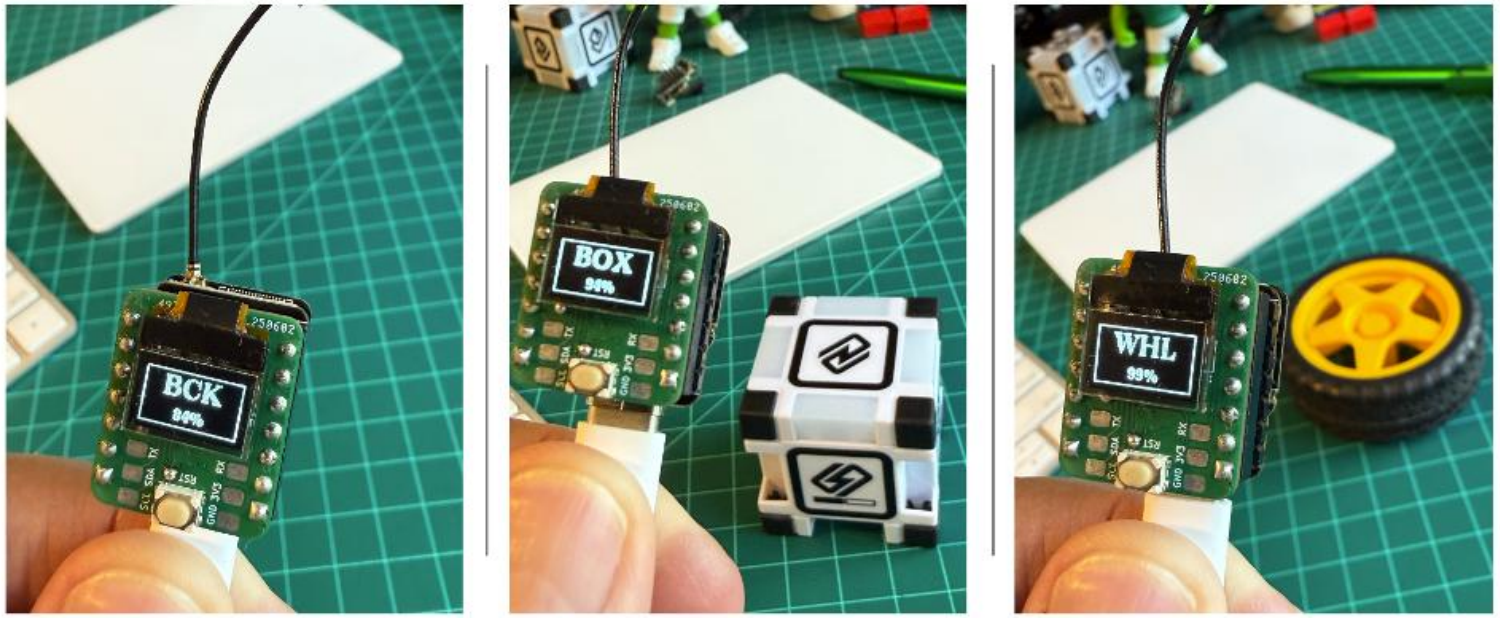
Deploy the model to edge device and implement the post-processing functions.





02 Project Example

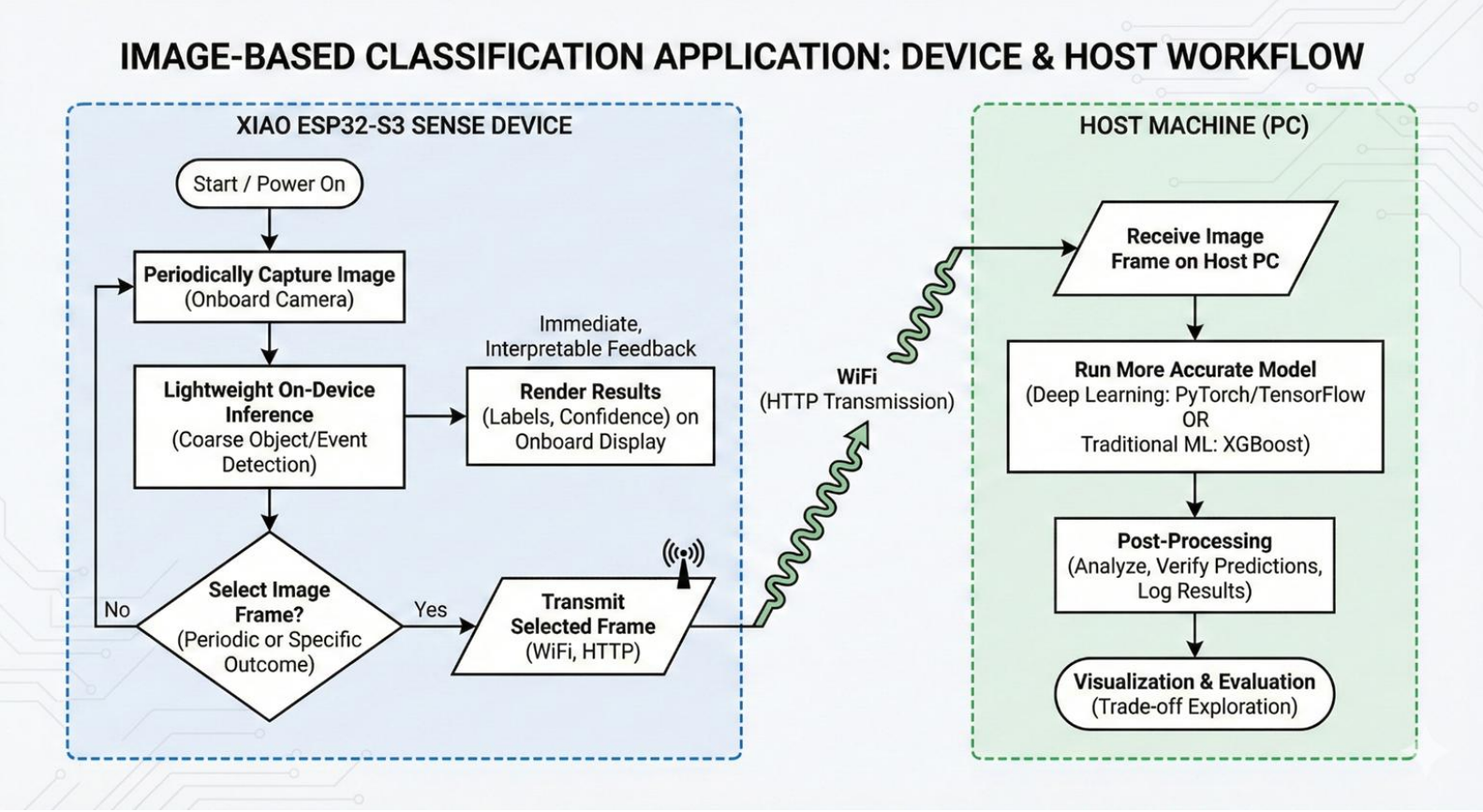
Test the performance of your model.



https://github.com/Mjrovai/XIAO-ESP32S3-Sense/blob/main/XIAOML_Kit_code/XIAOML-Kit-Img_Class_OLED_Gen/XIAOML-Kit-Img_Class_OLED_Gen.ino

02 Project Example

Advanced post-processing and IoT



02 Project Example

ESP32S3 Live Camera

Transmit captured image via WiFi

main.py > ...

```
4   from io import BytesIO
5
6   import torch
7   from transformers import AutoImageProcessor, AutoModelForImageClassification
8
9   # ================== 配置 ==================
10  ESP32_URL = "http://192.168.5.214/jpg"  # ← 改成你的 ESP32 IP
11  DEVICE = "cuda" if torch.cuda.is_available() else "cpu"
12
13  # ================== 加载 HF 模型 ==================
```

问题 4 输出 调试控制台 终端 端口

```
PS C:\Users\jyshe\OneDrive\桌面\ondevice> uv run .\main.py
  pencil sharpener                0.015
  mousetrap                       0.015

Top-5 Prediction:
  mouse, computer mouse           0.864
  cassette player                 0.024
  CD player                       0.007
  crash helmet                    0.006
  projector                       0.005

Top-5 Prediction:
  mouse, computer mouse           0.910
  cassette player                 0.009
  loudspeaker, speaker, speaker unit, loudspeaker system, speaker system 0.005
  projector                       0.004
  crash helmet                    0.003

Top-5 Prediction:
  mouse, computer mouse           0.851
  cassette player                 0.022
  space heater                    0.009
  loudspeaker, speaker, speaker unit, loudspeaker system, speaker system 0.009
  projector                       0.006

Top-5 Prediction:
  mouse, computer mouse           0.582
  cassette player                 0.043
```

Deploy a stronger model on PC

02 Project Bonus

- **Multi-modality Support:** Design a system that fuses information from more than one sensor simultaneously (e.g., combining Vision and Audio, or Vision and IMU).
- **Hardware Expansion:** Add and integrate new components (sensors, actuators, or servos) to the existing toolkit to enable novel functions.
- **Original Design:** Demonstrate a unique application scenario or system design that significantly differs from the standard course examples.

02 Important Dates

- [Mar 05] Project proposal
- [Mar 26] Mid-term project report
- [Apr 16] Demo presentation
- [Apr 16] Final report

Good luck with your on-device AI journey!

THANK YOU

01 References

https://mlsysbook.ai/book/

https://mlsysbook.ai/kits/

https://sensecraft.seeed.cc/ai/home

https://www.edgeimpulse.com/

https://github.com/Mjrovai/XIAO-ESP32S3-Sense/tree/main/XIAOML_Kit_code