TensorFlow Lite

The professional course

TensorFlow Lite

Week 2

Agenda

1. How to use the Android Profiler

   a. Overview of Android Profiler

   b. CPU profiler

   c. Memory profiler

   d. Energy profiler

   e. References

How to use the Android Profiler

Overview of Android Profiler
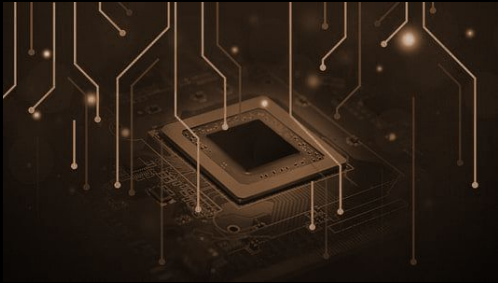
When an app is considered to have poor performance:

- It responds slowly

- Presents freezes and crashes

- Consumes a lot of power

Overview of Android Profiler

How to avoid these performance problems?

- Efficient usage of **CPU**, **memory,** and **battery**

Overview of Android Profiler

The Android Profiler tool provide real-time monitoring to help you to understand how your app uses CPU, memory, network, and battery (energy) resources.

Hands-On

Hands-on Project

Steps:

1. Overview
2. CPU profiler
3. Memory profiler
4. Battery profiler

How to use the official TF lite benchmark app

Agenda Part 2

2. How use the official TF lite benchmark app

    a. Overview

    b. Installing the benchmark app

    c. Running the app

    d. Filtering the results

    e. References

TF lite benchmark app

The TensorFlow lite benchmark tools currently measure and calculate statistics for the following important performance:

- Initialization time
- Inference time of warmup state
- Model inference time of steady-state
- Memory usage during initialization time
- Overall memory usage

Downloading and installing the benchmark app

Before running the benchmark app, install the app and push the model file to the device as follows:

- **Install the app:** adb install −r −d −g android_aarch64_benchmark_model.apk
- **Push the model:** adb push your_model.tflite /data/local/tmp

Running the app

After installed, we can run the application:

- **Running**: adb shell am start -S \ -n org.tensorflow.lite.benchmark/.BenchmarkModelActivity \ --es args

  '"--graph=/data/local/tmp/your_model.tflite \ --num_threads=1"'

- **Required params:**
  - **graph**: Path to the TFLite model

- **Optional params ***
  - **num_threads**: The number of threads to use for running TFLite interpreter.
  - **use_gpu**: Use GPU delegate
  - **num_nnapi**: Use NNAPI delegate
  - **num_xnnpack**: Use XNNPACK delegate

* Some of these options may not have available or have no effect, depending device you are using

Filtering the results

After install and run the app, we can see the results:

- **Filter results**: adb logcat | grep "Average inference"

Hands-On

Hands-on Project

Steps:

1. Download and install the benchmark app

2. Run the apk

3. Filter and analyse the results

Wrap-Up

Wrap-Up

During this week we learned:

1. The importance of app trace

2. How to see and analyze the hardware resources usage

3. How to measure the model inference time

References

1. https://developer.android.com/studio/profile/android-profiler

2. https://www.tensorflow.org/lite/performance/measurement#benchmark_tools