



TENSORFLOW LITE

NOTARO FABIO

DESCRIZIONE E STORIA

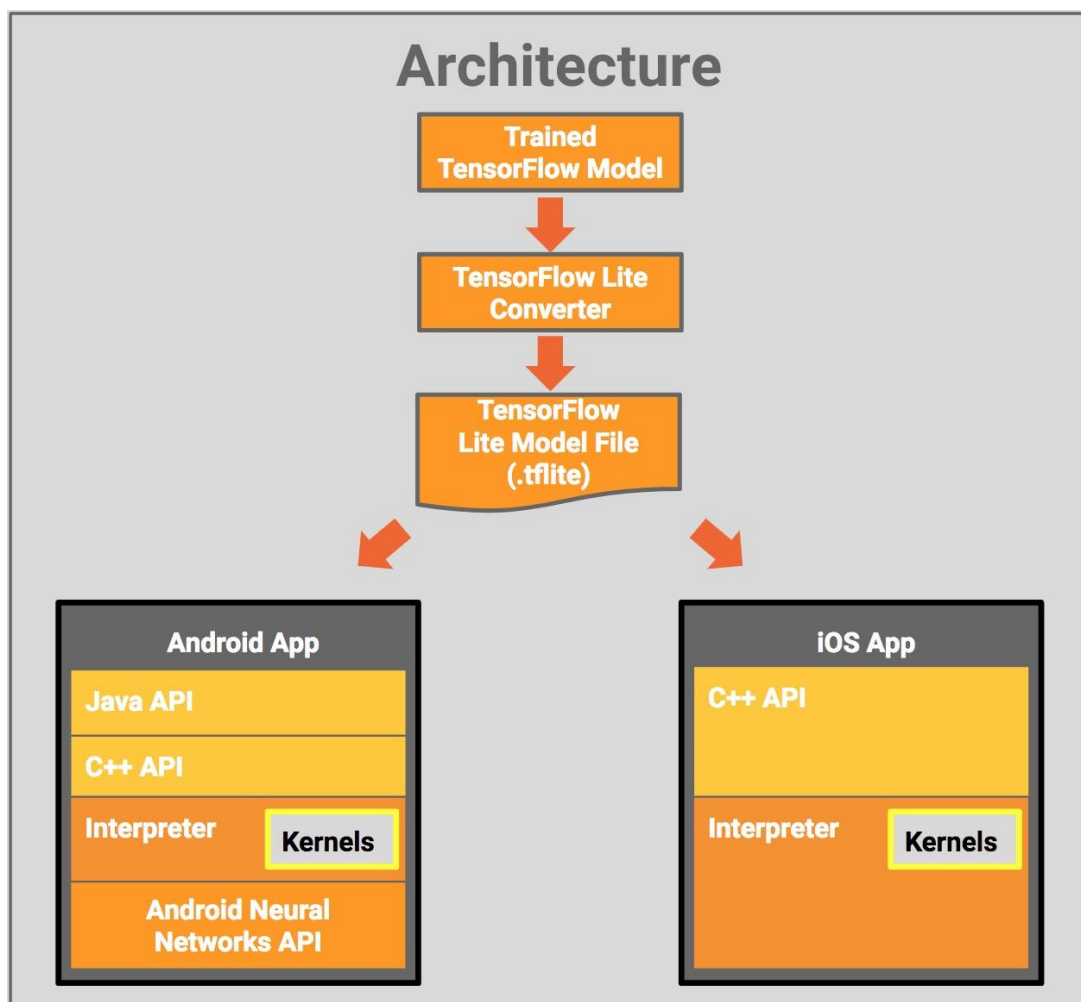
TensorFlow Lite è una libreria di machine learning open source sviluppata da Google per consentire l'esecuzione di modelli di intelligenza artificiale su dispositivi mobili, microcontroller e altri dispositivi a risorse limitate.

- E' stato sviluppato da un team di ingegneri di Google
- Annunciato per la prima volta nel maggio 2017
- Sviluppo e versioning incrementale e sempre più universale.

FUNZIONALITA'

- Utilizzo di diversi tipi di modelli di ML
- Compressione dei modelli al fine di ridurre la dimensione senza comprometterne le prestazioni
- Ottimizzazione delle prestazioni tramite parallelizzazione delle operazioni e l'uso di istruzioni SIMD
- Supporto per l'accelerazione hardware (se disponibile nel dispositivo) ad esempio tramite GPU
- Integrazione con TensorFlow tramite un'interfaccia semplice per convertire modelli addestrati con TensorFlow nella sua rappresentazione compatibile, nota come TensorFlow Lite FlatBuffer
- Ulteriori strumenti di conversione e quantizzazione
- Supporto per diversi linguaggi di programmazione, tra cui C++, Java, Python e Swift.

ARCHITETTURA



Esistono poi:

- librerie di supporto che forniscono API per caricare i modelli, gestire l'input e l'output, gestire le risorse di memoria...
- strumenti di ottimizzazione che consentono di ridurre le dimensioni dei modelli e migliorare le prestazioni delle inferenze.

IL PROCESSO DI CONVERSIONE

La conversione è il passaggio con cui un modello di machine learning addestrato con TensorFlow standard viene trasformato in un formato compatibile con TensorFlow Lite. Questo processo prevede:

- quantizzazione per ridurre le dimensioni e la precisione dei pesi del modello da float a numeri interi a 8 bit o meno
- ottimizzazione per ridurre la complessità computazionale del modello e migliorare le prestazioni
- conversione in formato FlatBuffer (binario)
- salvataggio del modello TensorFlow Lite su un file pronto per essere integrato nelle applicazioni mobile.

LE DIFFERENZE CON TENSORFLOW STANDARD

- Dimensioni dei modelli ridotte
- Efficienza computazionale ottimizzata
- Facilità di integrazione nelle applicazioni mobile
- Funzionalità specifiche per dispositivi mobile, ad esempio la segmentazione delle risorse disponibili.

VANTAGGI E SVANTAGGI

VANTAGGI	SVANTAGGI
Efficienza	Limitazione delle funzionalità
Velocità	Complessità di conversione dei modelli
Portabilità	Limitazioni di memoria e potenza di calcolo
Facilità di integrazione	Dipendenza dal sistema operativo e dalla piattaforma
Estensibilità	Inadatto a modelli ad elevata complessità

CODICE DI INTEGRAZIONE

```
dependencies {  
    implementation 'org.tensorflow:tensorflow-lite:2.5.0'  
}  
  
// Il modello deve trovarsi nella directory app/src/main/assets.  
val tflite = Interpreter(assets.openFd("nome_del_modello.tflite")) // Carica il modello TensorFlow Lite.  
// Esegui l'inferenza utilizzando il modello  
val inputBuffer = // Prepara i dati di input per il modello.  
val outputBuffer = // Prepara il buffer per i risultati di output.  
tflite.run(inputBuffer, outputBuffer)
```

Ulteriori informazioni su <https://www.tensorflow.org/lite/guide/android>.

ESEMPI DI SUCCESSO

App di
riconoscimento
immagini



App di traduzione in
tempo reale



App di assistenza
virtuale



App di controllo
gestuale



App di diagnostica
medica





GRAZIE DELL'
ATTENZIONE

NOTARO FABIO