

Luta dig tillbaka och låt

Här är Texas Instruments verktyg för att implementera maskininlärning på dess edge-processorer

Av Manisha Agrawal, TI Processors

Manisha Agrawal har tillbringat merparten av sin karriär på TI i Dallas, idag som produktmarknadschef, och på sistone med ansvar bland annat för marknadsföring och affärsutveckling kring AI-tillämpningar för processorfamiljen Jacinto inom edge och robotik. Manisha Agrawal läste till elektroingenjör på IIT Kanpur i Indien.



Figur 1. Arbetsflöde för utveckling av edge AI-applikationer.

Du vill ha en kort time-to-market, men det kan ta väldigt mycket tid att konstruera ett energieffektivt embeddedsystem som använder artificiell intelligens (AI) om du inte hittar några bra verktyg hos din processortillverkare. Det finns utmaningar i själva AI-algoritmen med att välja rätt deep learning-modell och att träna och optimera den att uppfylla kraven på prestanda och noggrannhet. Dessutom behöver du lära dig att använda de specifika verktyg som används för att implementera modellen på just den edgeprocessor du valt som målplattform.

Om det är ett djupt neuronät (DNN) du ska utveckla, för maskinvaruaccelererad inferens, så kan TI hjälpa dig hela vägen: från val av modell till implementering med verktyg, mjukvara och tjänster – allt gratis och med god nytta i varje steg, som visas i figur 1.

Här ska vi gå igenom stegen: hur du väljer modell, tränar den och implementerar den sömlöst i en TI-processor utan alltför mycket handpåläggning eller manuell programmering.

Steg 1: Val av modell

Den första uppgiften när du utvecklar ett edge AI-system är att välja rätt DNN-modell med avseende på prestanda, precision och strömförbrukning. Här kan du på Github hitta verktyget TI Edge AI Model Zoo som hjälper dig att snabba på den processen.

Model Zoo är en omfattande samling populära DNN-modeller i öppen källkod för Tensorflow, Pytorch och MX Net. Modellerna är tränade på ett öppet dataset och optimerade för att kunna köras effektivt på TI-processorer för AI-edge. TI uppdaterar regelbundet Model Zoo med de senaste modellerna från open source-världen såväl som med TI-design-

nade modeller för att ge dig ett brett urval av modeller för olika prestanda och precision.

Det finns hundratals modeller i Model Zoo. Med hjälp av TIs modellvalsverktyg (se figur 2) kan du snabbt jämföra och hitta rätt modell för din uppgift. Du kan jämföra statistik för prestanda som inferensflöde, latens, precision och double-data-rate-bandbredd utan att skriva en enda kodrad.

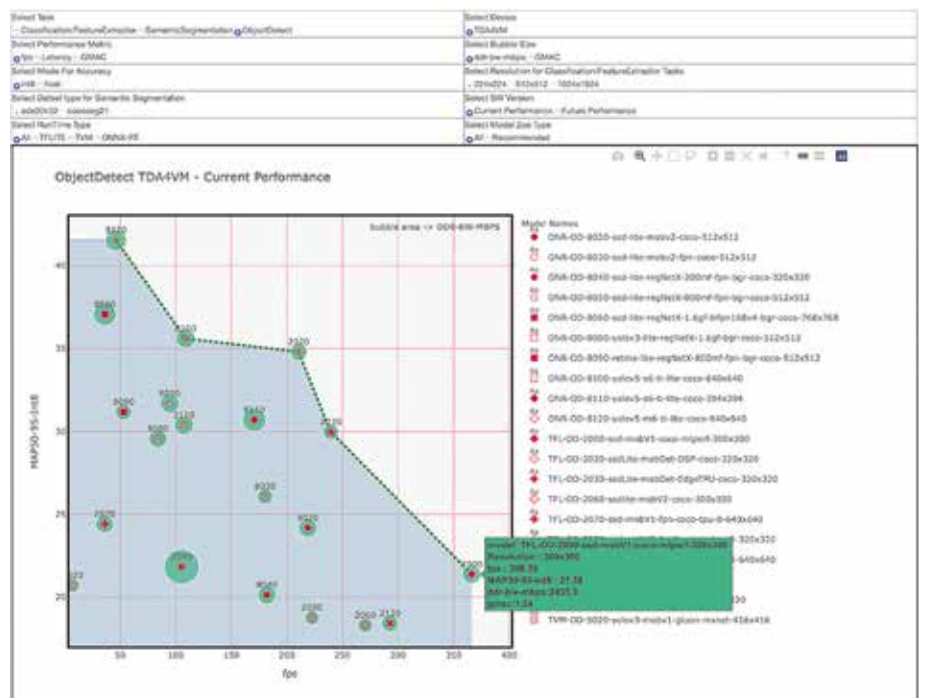
Steg 2: Träna och anpassa din modell

Efter att ha valt modell är nästa steg att träna och trimma den för att optimera prestanda

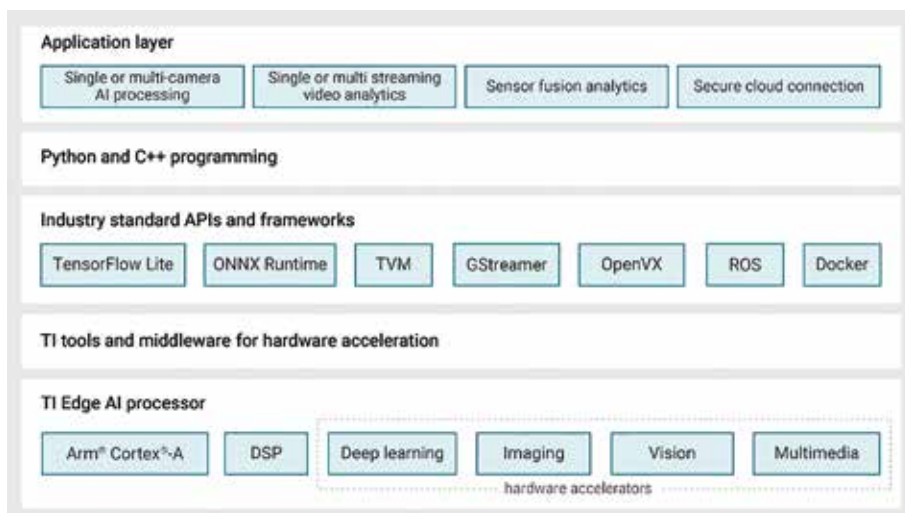
och precision för TI-processorer. Med hjälp av vår programvaruarkitektur och utvecklingsmiljö kan du träna modellen var du än befinner dig.

När du väljer en modell från TI Model Zoo kan du med hjälp av olika skript träna och transferera modeller för dina dataset, utan att behöva lägga tid på att träna från noll eller bygga modeller för hand. Du optimerar fram dina egna DNN-modeller med hjälp av träningskript, ramverkstillägg och verktyg som bland annat kan anpassa sig till valet av kvantisering.

Figur 2. TIs modellvalsverktyg.



TI bygga din AI



Figur 3. SDK med Linux för edge AI-komponenter.

Steg 3: Utvärdera modellprestanda

Utvärdering av modellprestanda på faktisk maskinvara krävs innan du utvecklar en edge AI-applikation.

TI:s flexibla programvaruarkitektur och utvecklingsmiljö gör att du kan träna din modell var som helst och kompilera och implementera den på TI-maskinvara med bara några rader kod. Använd de Python- eller C++-programgränssnitt du själv föredrar oavsett om de är för TensorFlow Lite, ONNX eller Neo AI DLR.

I backenden på dessa standardmotorer kan du med hjälp av våra TIDL-verktyg (TI Deep Learning) kompilera din modell för olika TI-hårdvara och ladda kompilerade grafer och subgrafer i DNN-acceleratorn. På det viset kan du krama inferensprestanda på toppnivå ur processorn – allt utan handpåläggning.

Under kompileringssteget kan du med hjälp av kvantiseringsverktyg för post-trä-

ning automatiskt konvertera flyttal till fixtal. Ett verktyg gör det möjligt att växla mellan olika precision (8 eller 16 bitar) i olika skikt. Du specificerar hur du vill ha det i en konfigurationsfil. Det här ger dig flexibilitet att anpassa en modell för högsta prestanda och precision.

Processen skiljer sig åt mellan olika modeller. TI Edge AI Benchmark Tool, som också den finns på GitHub, hjälper dig matcha funktionalitet sömlöst för modeller från TI Model Zoo och fungerar som en referens för skräddarsydda modeller.

Det finns två sätt att utvärdera modellprestanda på en TI-processor: TDA4VM Starter Kit Evaluation Module (EVM) och TI Edge AI Cloud som är en gratis onlinetjänst som ger fjärråtkomst till TDA4VM EVM för utvärdering av inferensprestanda för DNN-inlärning.

Det finns ett flertal exempelskript för olika kombinationer av uppgifter och runtime-

motorer som kan hjälpa dig programmera, implementera och därefter exekvera accelererad inferens på TI-maskinvara på mindre än fem minuter, samtidigt som du kan plocka ut benchmarks.

Steg 4: Utveckla edge AI-applikationer

Du kan använda API:er som är industristandard eller öppen källkod, som Linux, för att implementera din modell på TI-maskinvara. Att implementera din DNN-modell på en hårdvaruaccelerator är dock bara en bit av pusslet.

För att hjälpa dig att snabbt bygga upp en effektiv edge AI-applikation har TI börjat använda GStreamer-ramverket. Med GStreamer plug-in:er som körs med Arm-processorer som värd-cpu kan du automatiskt ta en signalkedja för en beräkningsintensiv uppgift och få den accelererad i en dedikerad hårdvaruaccelerator eller en DSP-kärna.

Figur 3 visar mjukvarustacken och komponenterna i processorn SDK med Linux för edge AI.

Slutsatser

Om du är nybörjare och känner dig överväldigad av de verktyg vi har pratat om i den här artikeln så kan du vara lugn – du behöver inte vara expert på AI och utvecklar och implementera en AI-modell, eller bygga en AI-applikation.

Med TI Edge AI Academy kan du lära dig grunderna inom AI och bygga upp en förståelse för AI-system och mjukvaruprogrammering i klassrumsmiljö med quiz där du själv bestämmer takten.

Labben introducerar steg-för-steg-kodning för att bygga en Hello World-AI-applikation. Du använder en avancerad tillämpning – en komplett signalkedja från kamera till display – för att utveckla en AI-applikation i din egen takt. ■

Ses vi igen?



Prenumerera gratis.

Du får månadsmagasinet på www.etn.se/pren

