Val av hårdvara

## Mål

* Vi vill undersöka olika typer av hårdvara som är lämpliga att implementera neurala nätverk på.
* Vi vill hitta den lösning som ger bäst balans mellan prestanda, enkelhet att implementera, och flexibilitet i implementationen.
* Vi är speciellt intresserade av hårdvara som passar för inference på lågenergi-hårdvara, som kan användas ”on the edge”.

## Förslag

Förslag på olika grupper av plattformar:

|  |  |  |
| --- | --- | --- |
| FPGA | GPU | ASIC |
| * Kan programmeras till det man önskar * Arbetar både parallellt och sekventiellt * Flexibel * Lågnivåspråk krävs (HDL) | * Många kärnor, kan jobba parallelt och är bättre för matrismultiplikation * Vanligast att använda GPU för ML * Väl testat | * Speciellt framtagen för att vara snabb på att lösa sin uppgift * Oflexibel * Bör alltid vara snabbast och mest energieffektiv |

Förslag på specifik hårdvara i varje grupp:

|  |  |  |
| --- | --- | --- |
| Arty Z7-20 | NVIDIA Jetson Nano | Google Coral dev board |
| * CPU: Cortex-A9 dual-core 650MHz * RAM: 512 MB * FPGA: XC7Z020 (jämförbar med Artix-7)   -Slices: 13,300  -LUTs: 53,200  -Flop-Flops: 106,400  -Block RAM: 630 kB | * CPU: Quad-core ARM A57 * RAM: 4 GB * GPU: 128 core NVIDIA Maxwell | * CPU: Cortex-A53 quad-core 1,5 GHz * RAM: 1 GB * TPU: Google Edge TPU, en ASIC för att köra NN ”on the edge” |
| * Har ett SoC med CPU och FPGA * Kan både använda frameworks som översätter till HDL och implementera HDL direkt * Kan bestämma vilka delar som ska exekveras i mjukvara (CPU) och i hårdvara (FPGA) | * Utvecklad för GPU-baserade ML-projekt. * Kör en version av Ubuntu = kan köra alla frameworks, inkl GPU-baserade. | * SoM med CPU och TPU (designad för inference). * Endast kompatibel med Tensorflow Lite-modeller |

## Alternativ

Alternativ till Arty Z7 skulle vara att köra på en ren FPGA, istället för att som med Arty ha FPGA kopplad till en CPU. Den FPGA som sitter i Arty är nästintill identisk med Artix-7, en FPGA också från Xilinx. Skillnaden är att vissa pinnar på Artix som används för konfigurering/bootning nu är kopplade till CPUn. Det gör att på Arty kan inte FPGAn jobba helt fristående, utan måste bootas med hjälp av CPUn. När bootning är klart fungerar dock FPGAn på Arty som en fristående FPGA. Skillnaden i bootningsprocess kommer inte att påverka ML-prestanda. Fördelen med CPU känns då större, eftersom att man kan använda färdiga frameworks och andra språk (t.ex. python/C) vilket gör implementeringen lättare.

Alternativ till Jetson Nano skulle vara en Raspberry Pi. Den stora fördelen med Jetson är att den har en mycket mer kraftfull GPU än RP, och vi kan därför bättre testa hur implementeringen av ML på en GPU fungerar.

## Mjukvara

Arty Z7 kan köpas med en voucher för SDSoC, vilket möjliggör att skriva i C/C++ och att använda färdiga libraries för både mjuk- och hårdvarukomponenter. För att skriva i HDL kan Vivado WebPack användas, vilket är gratis att ladda ner.

NVIDIA Jetson Nano kör en version av Ubuntu som heter Jetson, vilket är gratis. Därefter fungerar Jetson som en egen dator, och man kan sen köra färdiga frameworks som TensorFlow eller andra CUDA-kompatibla frameworks.

## Inköp

* Arty Z7-20 inkl voucher för SDSoC
  + <https://www.digikey.se/product-detail/sv/digilent-inc/471-006/1286-1221-ND/9826690>
  + Kostnad 2148,55 kr ex. moms.
* NVIDIA Jetson Nano
  + <https://www.elfa.se/sv/nvidia-jetson-nanotm-utvecklingssats-seeed-studio-102110268/p/30143589?queryFromSuggest=true>
  + Kostnad 1016 kr ex. moms.
* Tillbehör
  + <https://www.elfa.se/sv/usb-kabel-usb-svart-valueline-vlcp60500b10/p/30046369?track=true>
    - Micro USB-kabel till Arty och Jetson. Kostnad 25,10 kr ex. moms.
  + [<https://www.elfa.se/sv/microsd-ultra-memory-card-32-gb-sandisk-sdsquar-032g-gn6ma/p/30110932?track=true>](https://www.dustin.se/product/5011039978/ultra-c10-uhs-u1-a1)
    - SD-kort till Jetson Nano. Kostnad 116 kr ex. moms.
  + <https://www.digikey.se/product-detail/sv/cui-inc/SWI25-5-E-P5/102-4200-ND/7070096>
    - Ström till Jetson. Kostnad 165,68 kr ex. moms.
  + <https://www.digikey.se/product-detail/sv/cui-inc/SMI36-12-V-P5/102-3591-ND/5415062>
    - Ström till Arty. Kostnad 248,71 kr ex. moms.

Total kostnad inköp: 3720,04 kr ex. moms.

## Eventuella senare inköp

* Google Coral dev board – om tid och intresse finns att testa denna senare
  + <https://www.elfa.se/sv/coral-dev-kort-seeed-studio-102110260/p/30143588?queryFromSuggest=true>
  + Kostnad 1526 kr ex. moms.
* Raspberry Pi 4/4 GB – om tid och intresse finns att använda denna som referens
  + <https://www.dustin.se/product/5011141621/4-model-b-4gb>
  + Kostnad 559 kr ex. moms.
* Kameror till korten för att kunna köra live. Något i stil med Raspberry pi-kamera/arduino shield-kamera.
* Material till att bygga någon slags låda, för att få till en ”slutprodukt”.