

Kennismaking met TinyML

groot is niet altijd beter

José Bagur (Guatemala)

Een van de snelst groeiende gebieden van deep learning is tiny machine learning (TinyML). TinyML is een geavanceerd gebied dat modellen voor machinaal leren onderbrengt in zuinige en goedkope computerapparatuur, zoals microcontrollers. In dit artikel wordt beschreven waarom TinyML bewijst dat groot niet altijd beter is.

Wat is TinyML?

TinyML is een deelgebied van machine learning (ML) dat zich richt op de ontwikkeling van modellen die in realtime kunnen worden uitgevoerd op low-power, goedkope embedded apparaten [1]. De ontwikkeling van TinyML-modellen volgt het typische ML-proces zoals weergegeven in **figuur 1**, met dit verschil dat de inferentie plaatsvindt op embedded apparaten in plaats van op traditionele computers of in cloud-gebaseerde diensten.

Gewoonlijk gebruiken TinyML-modellen gegevens die zijn verzameld van IoT-apparaten; deze gegevens worden gebruikt om een model te trainen in een service (meestal in de cloud) om kennispatronen uit de dataset halen. Deze patronen worden vervolgens

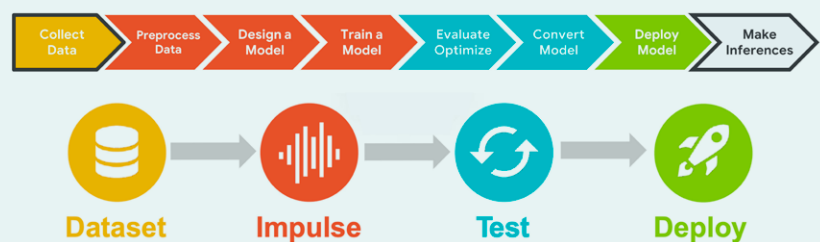
verpakt in een model dat rekening houdt met de capaciteiten van het embedded apparaat, zoals geheugen en rekenkracht [1]. Het resulterende model kan vervolgens worden ingezet in embedded apparaten die nieuwe sensorgegevens in realtime en ter plaatse evalueren, zonder gebruik te maken van externe bronnen zoals cloud-gebaseerde services. TinyML-toepassingen verbruiken

doorgaans weinig energie (meestal enkele mW); dit maakt het mogelijk dat batterijgevoede apparaten deel gaan uitmaken van het ML-universum (**figuur 2**).

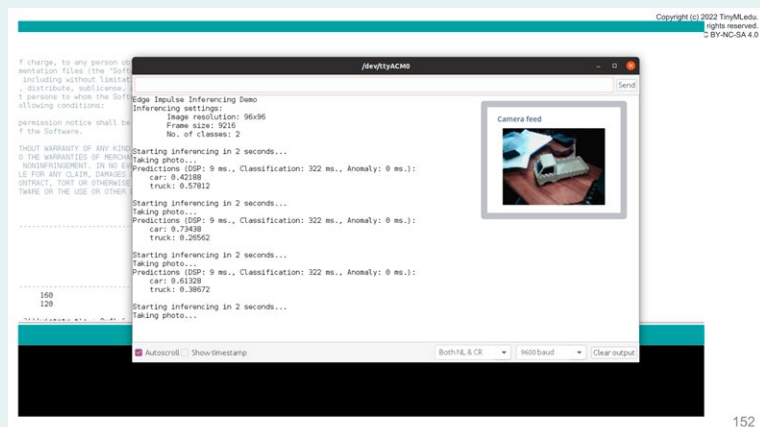
Belangrijkste kenmerken van TinyML

Nu we weten wat TinyML is, zetten we de belangrijkste kenmerken ervan op een rijtje:

The TinyML Workflow using Edge Impulse



Figuur 1. Typische TinyML-workflow (bron: TinyMLedu).



Figuur 2. Inferentieproces in een embedded systeem (bron: TinyMLedu).

- **Latentie:** aangezien de inferentie direct op het embedded apparaat plaatsvindt, is de latentie in TinyML-toepassingen gering.
- **Energieverbruik:** TinyML-modellen houden rekening met de beperkingen van embedded apparaten; ze draaien ook op apparaten met een laag stroomverbruik, zoals microcontrollers, waarvan het stroomverbruik meestal enkele mW bedraagt. Dit betekent dat batterijgevoede apparaten kunnen worden gebruikt voor TinyML-toepassingen.
- **Bandbreedte:** aangezien het model rechtstreeks in het embedded apparaat draait, hoeven de verzamelde gegevens niet naar een externe service te worden gestuurd; dit betekent dat er minder internetbandbreedte wordt gebruikt.
- **Privacy:** gegevens die in TinyML-modellen worden gebruikt, worden realtime en ter plaatse verzameld en geanalyseerd; gegevens worden op geen enkel moment verzonden of gedeeld met externe services.

TinyML-toepassingen en gebruiksvoorbeelden

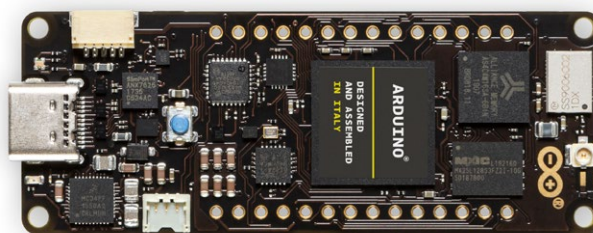
TinyML heeft het enorme potentieel om de bottleneck van IoT-toepassingen weg te nemen: gegevens. Aangezien inferentie lokaal plaatsvindt, maakt TinyML het tijdperk van het *Internet of Thinking Things* (IoT2) mogelijk, wat een eindeloze reeks van verbeterde en nieuwe toepassingen betekent.

Enkele voorbeelden van werkelijke toepassingen en gebruikssituaties van TinyML zijn:

- **Energiebesparing/optimalisatie:** het stroomverbruik kan binnen een TinyML-model drastisch worden verbeterd door piekvermogen te leveren bij een zo laag mogelijk verbruik. Dit kan met behulp van TinyML worden vertaald naar de elektrische of zelfs mechanische wereld.
- **Voorspelling van natuurrampen (in een vroeg stadium):** vroege stadia van natuurrampen zouden kunnen worden herkend om grote schade aan de bestaande infrastructuur te voorkomen met behulp van verschillende apparaten met een TinyML-model erin geïnjecteerd in een mesh-netwerk. Dit kan worden bereikt door het leren van een breed



Figuur 3. De Arduino Tiny Machine Learning Kit bevat een Nano 33 BLE Sense board.



Figuur 4. Arduino Portenta H7.

spectrum van signalen en patronen die kunnen optreden in gebieden met een verhoogd risico.

- **Herkennen van gezondheidsproblemen:** abnormale gezondheidsproblemen zouden eerder kunnen worden opgespoord om plotseling letsel of overlijden te voorkomen. Hartaandoeningen zouden bijvoorbeeld tijdig kunnen worden herkend zodat de betroffene een arts kan opzoeken. Het TinyML-model kan worden gerealiseerd in een draagbaar apparaat dat in noodgevallen verbinding kan maken met externe services.

Arduino en TinyML

Arduino biedt verschillende hardware-opties en softwarebibliotheken die gebruikt kunnen worden voor TinyML-toepassingen. Laten we twee prima boards bekijken die een goed uitgangspunt vormen voor het ontwikkelen van TinyML-toepassingen: de Arduino Nano 33 BLE Sense en de Arduino Portenta H7.

De Arduino Nano 33 BLE Sense (**figuur 3**) gebruikt de nRF52840 van Nordic Semiconductor; dit is een 32-bit ARM Cortex-M4F microcontroller die op 64 MHz draait, met 1 MB flash en 256 KB RAM [2]. De Nano 33 BLE Sense heeft verschillende sensoren aan boord die in veel TinyML-toepassingen van pas kunnen komen:

- versnellingsmeter, gyroscoop en magnetometer (LSM9DS1)
- microfoon (MP34DT05)
- gebaar, licht en nabijheid (APDS9960)
- barometrische druk (LPS22HB)
- temperatuur en luchtvochtigheid (HTS221)

De Arduino Portenta H7 (**figuur 4**) is een krachtig board voor industrieel gebruik, ontworpen voor veeleisende toepassingen. De Portenta H7 gebruikt de STM32H747-microcontroller van STMicroelectronics; deze microcontroller combineert een Cortex M7-kern op 480 MHz en een Cortex M4-kern op 240 MHz.

De Portenta H7 kan gelijktijdig high-level code en realtime-taken uitvoeren. Met de Portenta H7 kunnen we bijvoorbeeld gecompileerde Arduino-code draaien samen met gecompileerde MicroPython-code en beide kernen laten communiceren via het Remote Procedure Call-mechanisme (RPC) [3].

Qua software kunnen de Nano 33 BLE Sense en de Portenta H7 gebruik maken van het TensorFlow Lite software-framework voor het ontwikkelen van TinyML-modellen; de *Artificial Intelligence for Embedded Systems* (AlfES), ontwikkeld door het Fraunhofer Institute for Microelectronic Circuits and Systems, is ook een geweldig software-framework dat is geoptimaliseerd voor embedded systemen. Edge Impulse, een cloud-gebaseerde ML-service, wint ook aan populariteit in de community – het ondersteunt Arduino-boards zoals de Nano 33 BLE Sense en de Portenta H7.

Verdere resources

Het leren kennen van een techniek die nog in ontwikkeling is kan moeilijk zijn, maar voor TinyML zijn er uitstekende online-resources:

- Het boek "TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power" van Pete Warden en Daniel Situnayake [4] moet je gelezen hebben; het is een goed uitgangspunt voor het verkennen van het TinyML-universum.
- *The Professional Certificate in TinyML* [5] van Harvard University is beschikbaar op edX. Deze gratis online-specialisatie (vier cursussen) duikt dieper in het TinyML-universum.
- De cursus *Introduction to Embedded Machine* [6] van Edge Impulse is beschikbaar op Coursera. Deze gratis online-cursus geeft ook een breed overzicht van hoe ML werkt, hoe je TinyML-modellen kunt trainen met Edge

Impulse, en hoe je die modellen kunt inzetten op microcontrollers.

- Het *TinyML Open Education Initiative* (TinyMLedu) [7]. Dit initiatief is een internationale groep van academici en professionals uit de industrie die werken aan een betere wereldwijde toegang tot onderwijsmateriaal voor het geavanceerde gebied van TinyML.
- *TinyML for Developing Countries* (TinyML4D) [8]. Dit initiatief werkt aan de ontwikkeling van content voor een netwerk van onderzoekers en praktijkmensen dat zich richt op het verwerken van innovatieve oplossingen voor de unieke uitdagingen waarmee ontwikkelingslanden zich geconfronteerd zien.

Het IoT2-tijdperk

TinyML is een opkomende techniek rond ML-modellen die kunnen worden ingezet in kleine, goedkope en energiezuinige apparaten zoals microcontrollers. Met de veelzijdigheid van hardware- en software-tools zoals het Arduino-ecosysteem en software-frameworks zoals TensorFlow Lite en Edge Impulse komt het IoT2-tijdperk onder handbereik. ◀

220573-03

Over de auteur



José Bagur is docent en onderzoeker aan de Universidad del Valle de Guatemala (UVG). Hij studeerde Mechatronics Engineering aan de UVG voordat hij een master in IoT behaalde aan de Universiteit van Salamanca. Met een bijzondere interesse in ruimtevaart-gerelateerde projecten is José's onderzoek gericht op de ontwikkeling van open-source low-cost nanosatelliet-hardware. José werkt ook voor Arduino als content creator.

Vragen of opmerkingen?

Hebt u technische vragen of opmerkingen over dit artikel? Stuur een e-mail naar de redactie van Elektor via redactie@elektor.com.



Gerelateerde producten

Op zoek naar de belangrijkste producten die in dit artikel zijn genoemd? Arduino en Elektor hebben ze voor je klaargezet!

- **Arduino Tiny Machine Learning Kit (SKU 19943)**
www.elektor.nl/arduino-tiny-machine-learning-kit
- **Arduino Pro Nicla Vision (SKU 20152)**
www.elektor.nl/20152
- **Arduino Portenta H7 Development Board (SKU 19351)**
www.elektor.nl/19351

WEBLINKS

- [1] M. Zennaro, "TinyML: Applied AI for Development Challenges with Machine Learning in Developing Countries": <https://sdgs.un.org/sites/default/files/2022-05/2.1.3-9-Zennaro-TinyML.pdf>
- [2] Nano 33 BLE Sense: <https://docs.arduino.cc/hardware/nano-33-ble-sense>
- [3] Portenta H7: <https://docs.arduino.cc/hardware/portenta-h7>
- [4] P. Warden & D. Situnayake, TinyML (O'Reilly Media, 2019): www.oreilly.com/library/view/tinyml/9781492052036/
- [5] Tiny Machine Learning Open Education Initiative (TinyMLedu): <https://tinyml.seas.harvard.edu/>
- [6] TinyML4D: TinyML for Developing Countries: <http://tinymledu.org/4D>
- [7] Tiny Machine Learning (TinyML) Professional Certificate: www.edx.org/professional-certificate/harvardx-tiny-machine-learning
- [8] Inleiding tot Embedded Machine Learning: www.coursera.org/learn/introduction-toembedded-machine-learning