

computer vision on edge devices

TONI BADERTSCHER

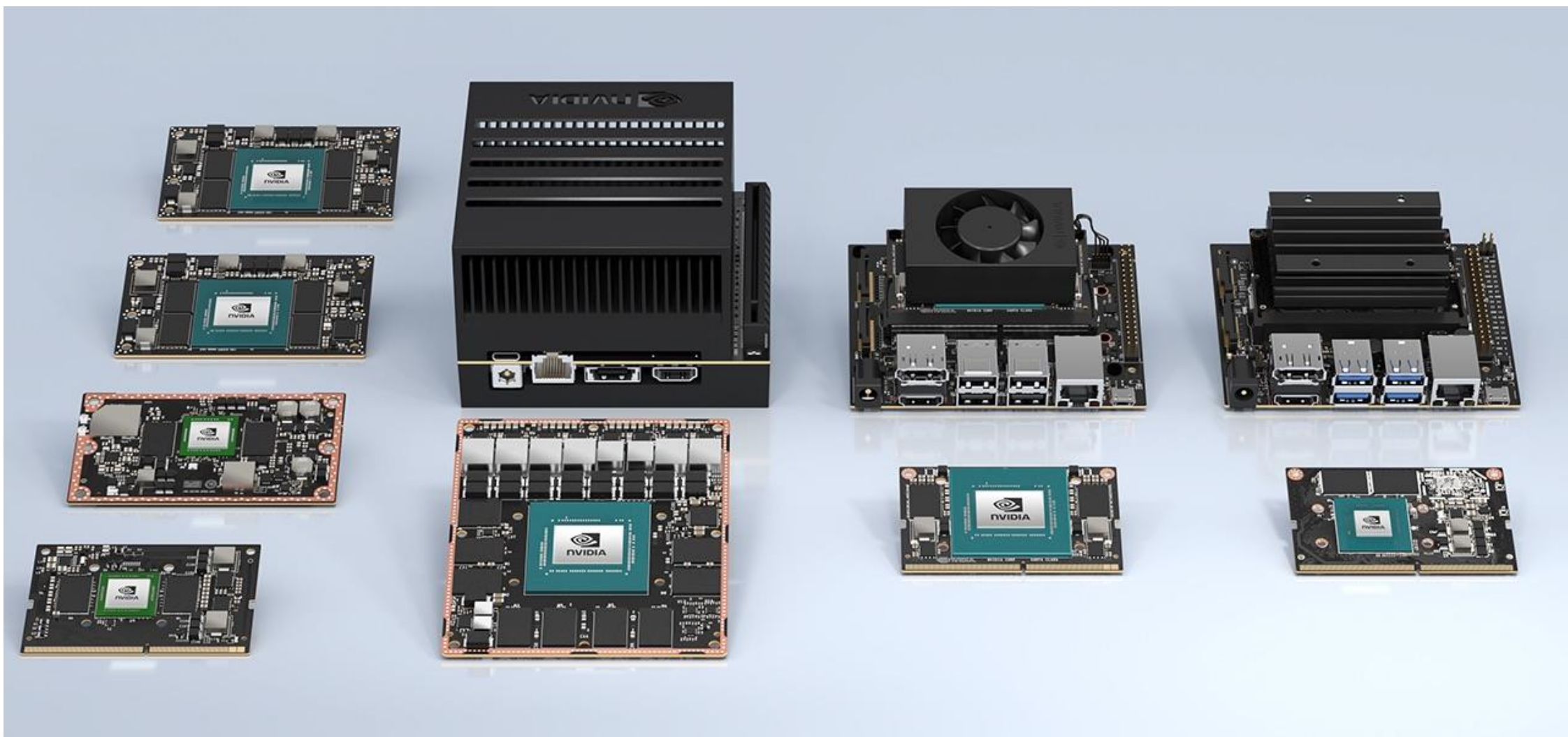
MACHINE LEARNING PROJEKT

Agenda

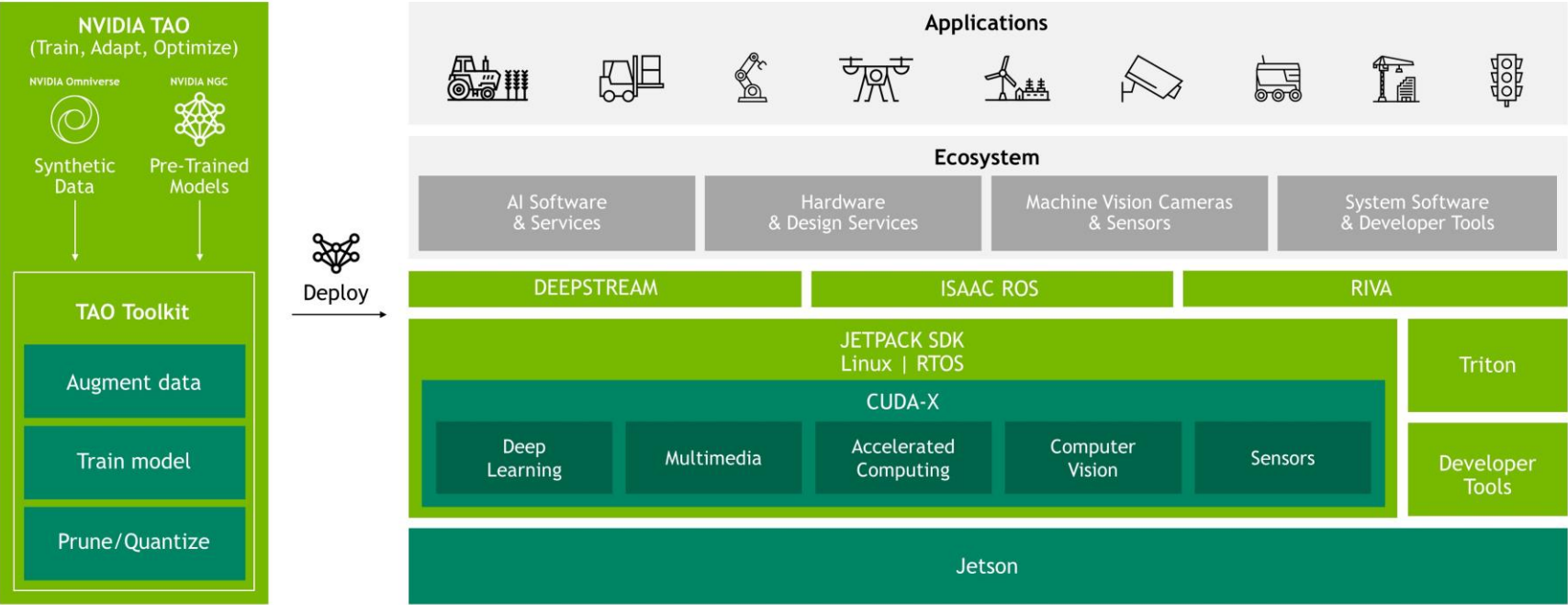
- Projektidee / Motivation
- Timeline
- Hardware setup, Entwicklungsumgebung
- Analyse, funktionale und nicht-funktionale Anforderungen
- MVP
- Erweiterungen, Verbesserungen
- Fazit, lessons learned

Jetson hardware & dev kits





JETSON SOFTWARE

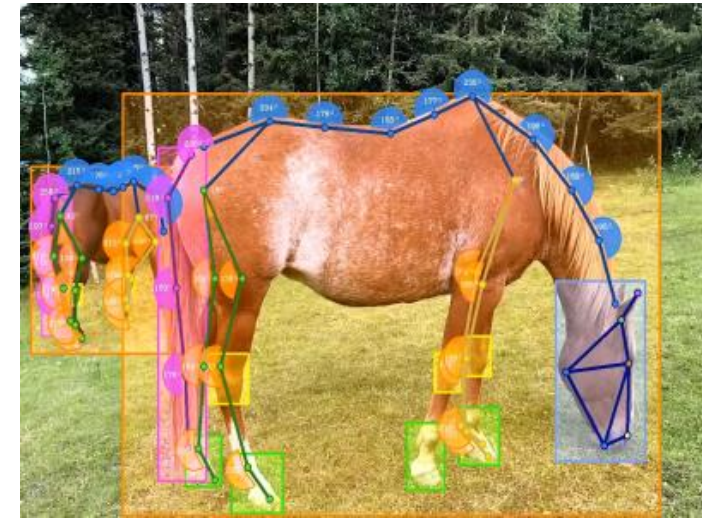
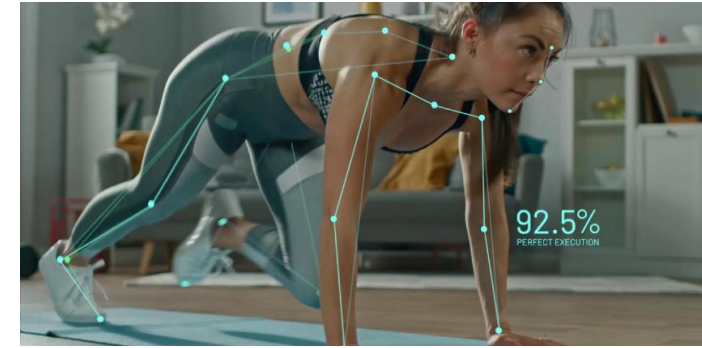




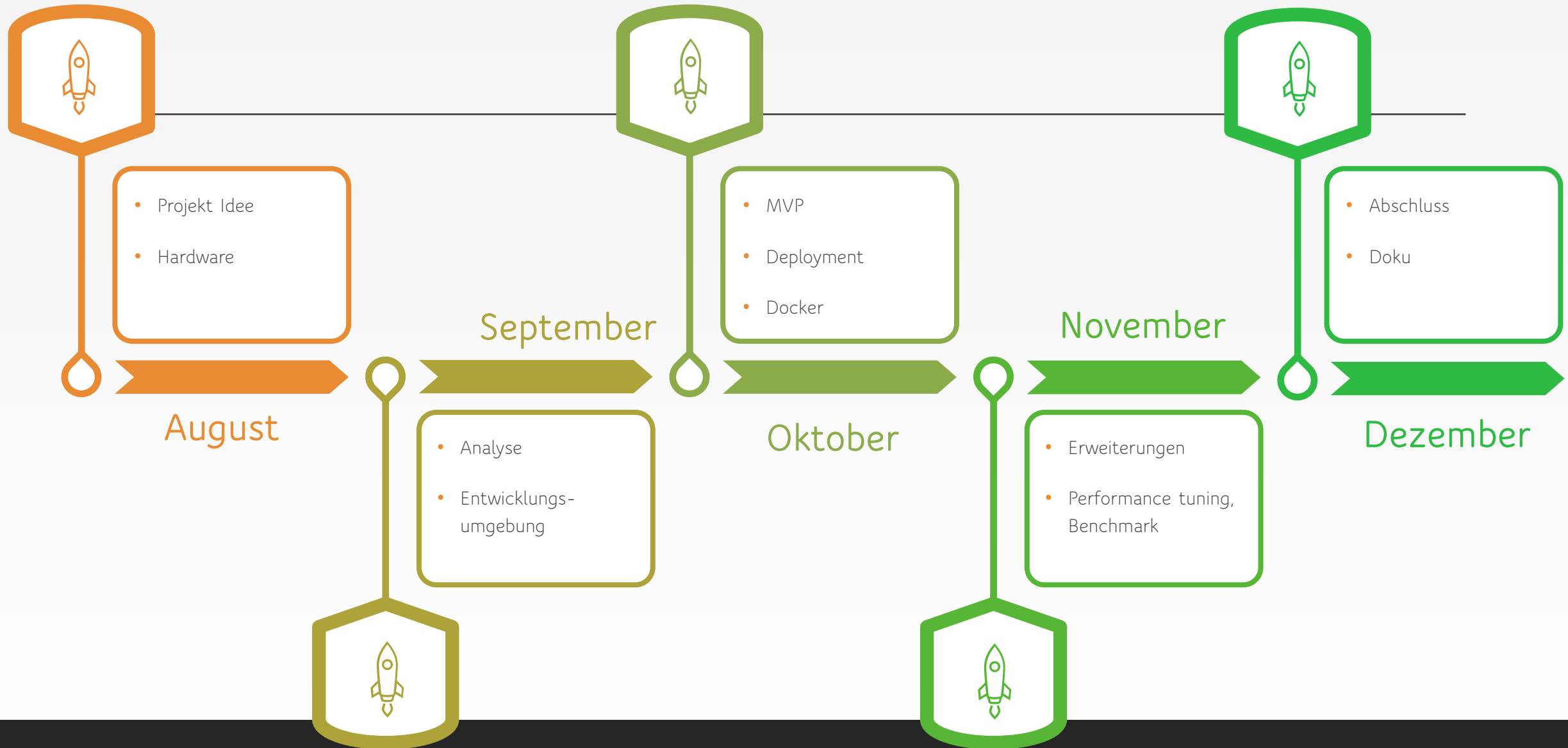
pose estimation / classification

Anwendungsbereiche

- human - robot interaction
- fitness tracker
- motion tracking for gaming
- avatar / metaverse
- livestock monitoring

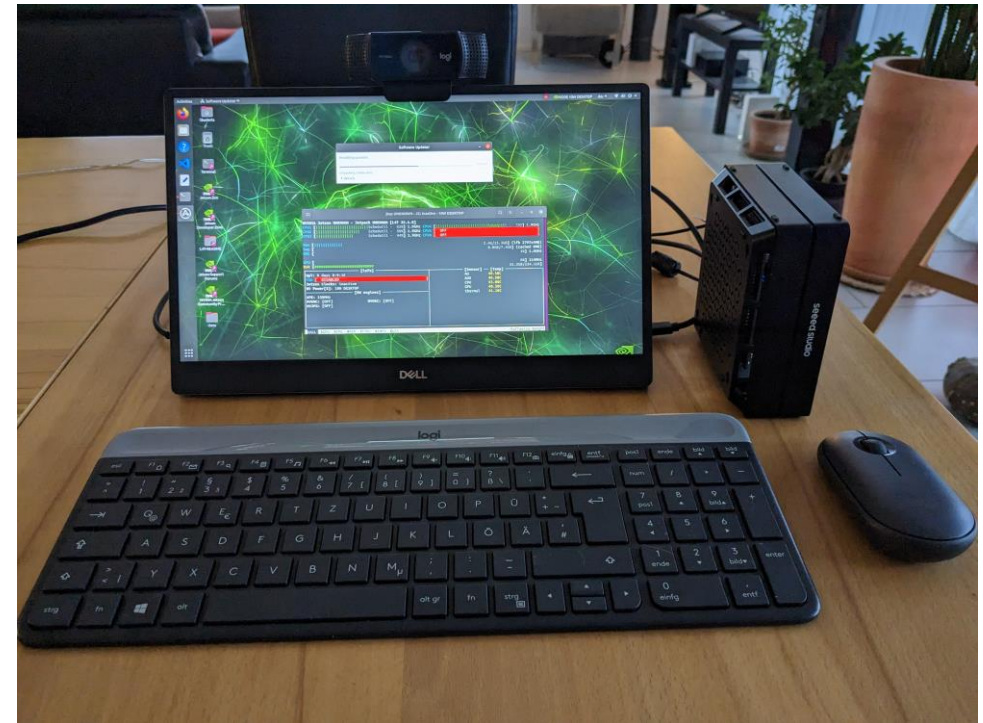


Projekt timeline

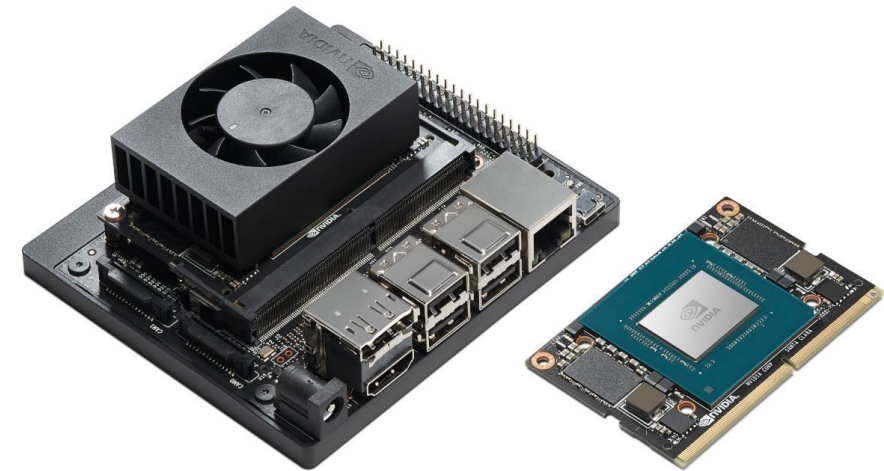
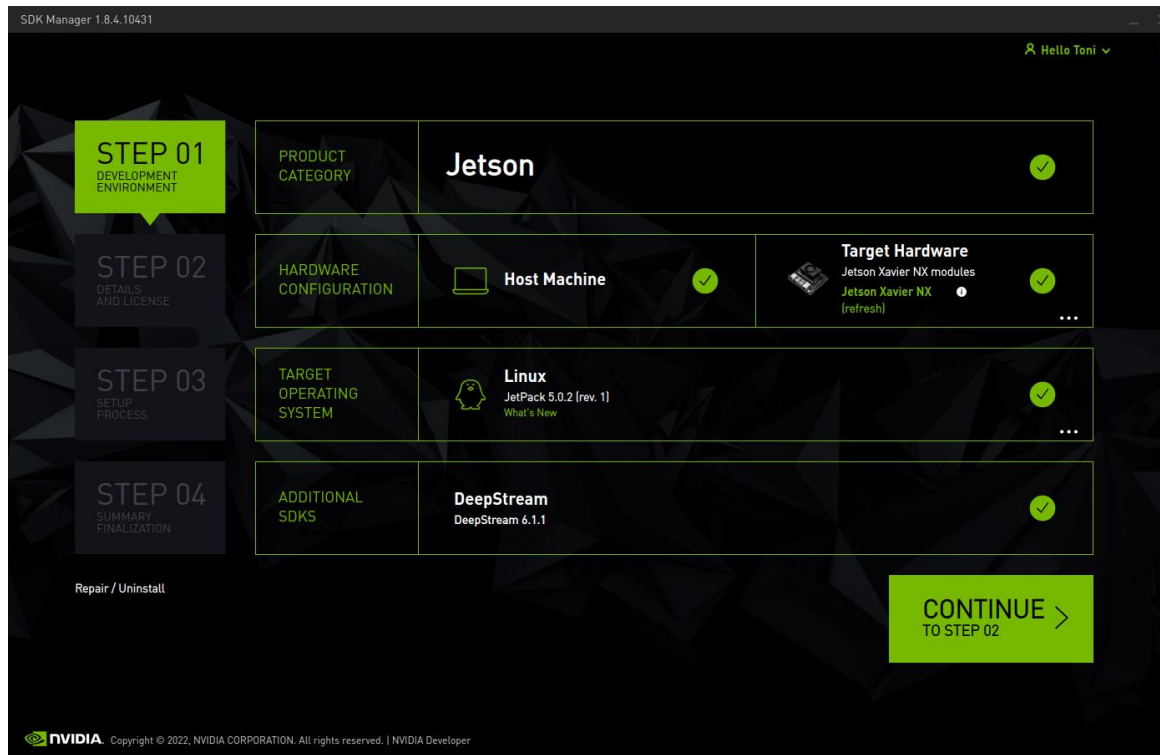


Hardware setup, tools

- Jetson Xavier NX, 16GB RAM, 256GB SSD, USB Webcam, RTSP camera
- Jetpack 5.02
- Ubuntu 20.04, ARM
- Cuda, cuDNN, TensorRT, Pytorch, Deepstream, TAO
- Docker, Nvidia NGC CLI
- X86 Laptop, 32GB RAM, RTX3060 6GB, Ubuntu 20.04



Jetpack, SDK Manager



Analyse

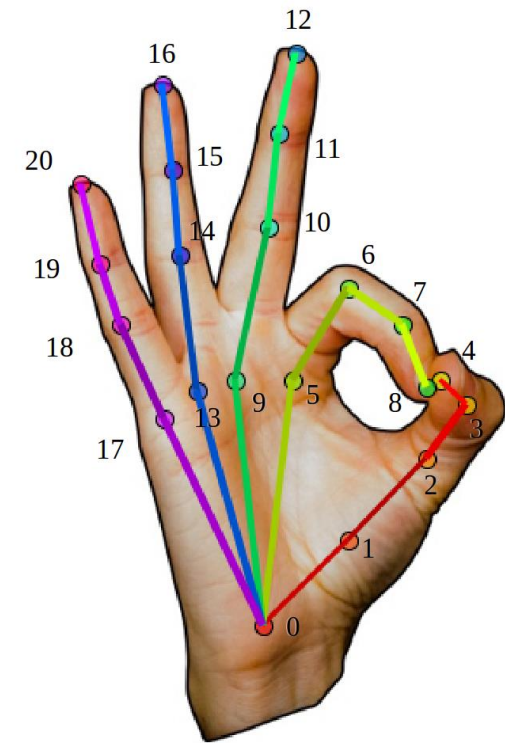
- Olga Chernytska, 2D Hand Pose Estimation-RGB, FreiHand dataset
- NVIDIA AI IOT: trt_pose, trt_pose_hand
- Google mediapipe (python)
- MMPose (OpenMMLab Pose Estimation Toolbox)
- YOLOv7
- Nvidia TAO, DeepStream, Triton

Anforderungen

- 2d oder 3d keypoint estimation auf der Basis von RGB input (webcam)
- Klassifizierung von Handzeichen
- single shot bottom up oder top down (detection model & pose estimation)
- single person (ein oder beidhändig)
- als Basis sollen PyTorch Modelle dienen, Erweiterbarkeit
- real-time, mindestens 5fps
- target platform: X86 und jetson (ARM)

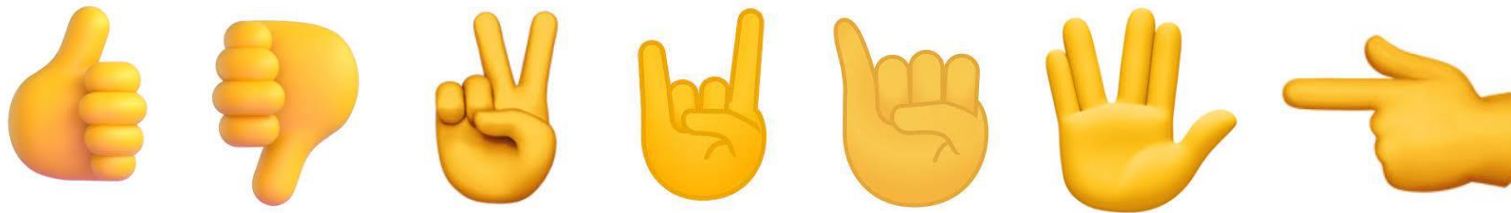
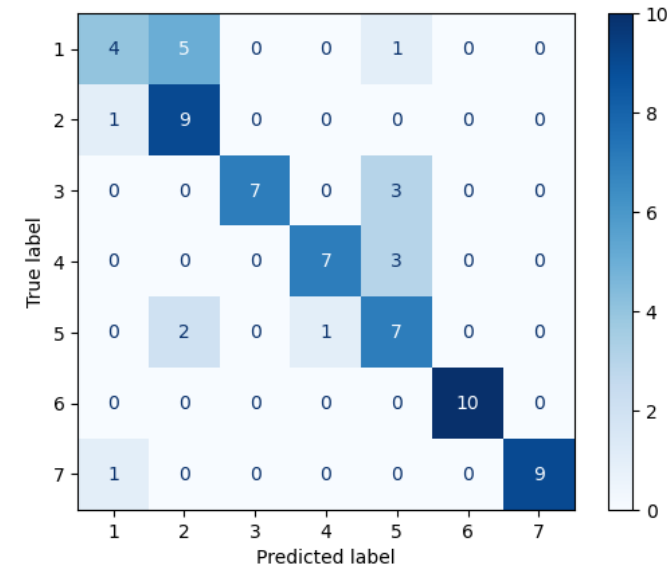
MVP

- basiert auf NVIDIA IOT demo (trt_pose / trt_pose_hand)
- resnet18 baseline
- single hand, 2d keypoint estimation
- scikit-learn SVM classifier
- torch2trt (PyTorch to TensorRT converter)



Classifier

- Datenset mit 7 Klassen (home made)
- 40 Bilder pro Klasse für Training
- 10 Bilder pro Klasse für Test
- svm accuracy: 0.76



Performance (Cuda, TensorRT)

- fp16 half-precision inference performance
- Xavier NX, 12 nm, 384-core NVIDIA Volta GPU, 48 Tensor Cores
- RTX 3060 notebook GPU, 8 nm, 3840-core Ampere GPU

	batch size 1	batch size 8
AMD Ryzen RTX3060 CPU	118.93 ms	714.60 ms
AMD Ryzen RTX3060 GPU	5.43 ms	30.56 ms
AMD Ryzen RTX3060 TRT	1.07 ms	16.46 ms
Jetson Xavier NX CPU	7534.95 ms	28720.30 ms
Jetson Xavier NX GPU	36.37 ms	343.61 ms
Jetson Xavier NX TRT	5.96 ms	24.80 ms

Ertweiterungen, SOT

- MMPose
- ViTPose: vision transformer for human pose estimation
- Facebook (Meta research) InterHand2.6M (3D)
- Nvidia 3d pose estimation
- Deepstream, TAO, Triton
- Nvidia transformer

Conclusion

- there is more than meets the eye
- vom research paper zu einer brauchbaren Anwendung ist ein langer Weg
- hardware und library Probleme kosten schnell viel Zeit
- frameworks haben oft eine Lernkurve. Dokumentation & support meist limitiert
- frameworks machen komplexe tasks einfacher. Vielfach sind aber einfache Sachen komplex