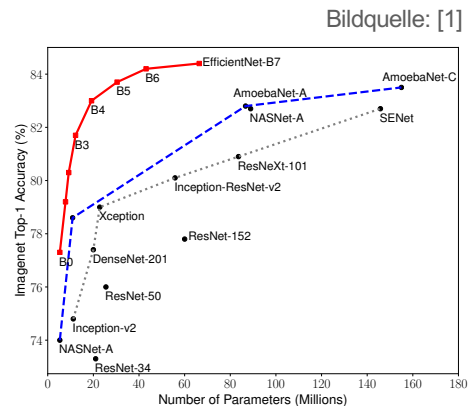


Masterarbeit

Thema: Analyse der Leistungsfähigkeit des EfficientNet für den Einsatz auf einer mobilen Plattform

Beschreibung:

Durch den Einsatz von Convolutional Neural Networks konnte die Leistung bildbasierter Klassifikatoren in den letzten Jahren immer weiter gesteigert werden. Inzwischen übertreffen aktuelle Deep-Learning-basierte Klassifikatoren sogar die menschliche Leistung in vielen Anwendungsbereichen. Ermöglicht wird eine solch hohe Klassifikationsgüte durch den Einsatz von Neuronalen Netzen mit bis zu 1000 Schichten. Für das Training und die Anwendung derart tiefer Neuronaler Netze bedarf es leistungsstarker GPUs oder gar speziell optimierter TPUs. Während diese Ressourcen in vielen Einsatzszenarien zur Verfügung stehen, ist die Rechenleistung auf mobilen Plattformen, wie etwa den im Fachgebiet NIKR eingesetzten Robotern, limitiert. In der Literatur gibt es daher eine Vielzahl von Ansätzen, in denen bekannte Architekturen speziell für den mobilen Einsatz optimiert werden. Ziel dieser Masterarbeit ist es, die Leistungsfähigkeit der vielversprechenden Architektur „EfficientNet“ [1, 2] und ausgewählter Modifikationen [3, 4] zu analysieren.



Detaillierte Aufgabenstellung:

- Systematische Aufarbeitung des State of the Art zum EfficientNet [1, 2] und der ausgewählten Erweiterungen [3, 4]
- Implementierung in PyTorch in Anlehnung an [6]
- Überführung mittels ONNX zu NVIDIA TensorRT [5] für den Einsatz auf mobilen Robotern in Verbindung mit einem NVIDIA Jetson AGX Xavier und Bestimmung von Inferenzzeiten
- Evaluierung der Modifikationen auf einem zugeschnittenen Datensatz (Implementierung, Training, Bewertung)
- Ausarbeitung von Präsentationen für den Eröffnungs-, und Abschlussvortrag
- Anfertigen der Masterarbeit entsprechend der Vorgaben des FG NI&KR

Notwendige Voraussetzungen:

- Verständnis englischsprachiger Literatur
- Gutes mathematisches Verständnis und Erfahrung in der Programmierung mit Python

Literatur:

- [1] Tan et al.: [EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks](#), ICML, 2019.
 - [2] Tan et al.: [MnasNet: Platform-Aware Neural Architecture Search for Mobile](#), CVPR, 2019.
 - [3] Chen et al.: [Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution](#), ICCV 2019.
 - [4] Li et al.: [DABNet: Depth-wise Asymmetric Bottleneck for Real-time Semantic Segmentation](#), BMVC, 2019.
 - [5] NVIDIA TensorRT: <https://developer.nvidia.com/tensorrt> (online: 04.12.2019)
 - [6] TorchVision: <https://github.com/pytorch/vision/issues/980> (online: 04.12.2019)
- IEEE Recherchesystem www.ieeexplore.ieee.org (nur aus dem Uni-Netz bzw. via VPN)
 - Google Scholar scholar.google.com
 - Proceedings der relevanten Konferenzen (NIPS, ICML, ICLR, IJCNN, WCCI, ICANN, CVPR, ICCV, ECCV, BMVC, ICPR, ICIP, ...)

Betreuer: Daniel Seichter, M.Sc. (daniel.seichter@tu-ilmenau.de)
Betr. Hochschullehrer: Prof. Dr. H.M. Groß
Bearbeiter: Nina Sorokina