

Hauptseminar

Thema: Vortraining von Deep-Learning-Architekturen mittels Autoencodern

In der Anfangszeit des Deep Learning wurden Autoencoder bereits verwendet, um tiefe Architekturen vorzutrainieren. Durch geeignete Gewichtsinitialisierungen und Techniken, die den Gradientenfluss verbessern, wurde ein Vortraining überflüssig und die Deep-Learning-Architekturen wurden zunehmend vollständig überwacht trainiert. In den letzten Jahren rückte unüberwachtes Training aber wieder stärker in den Fokus, um auch große ungelabelte Datenbestände nutzen zu können. Dabei sind in aktuellen Forschungsarbeiten auch Autoencoder wieder in den Fokus gerückt. In [1] wurde gezeigt, wie mit einem Masked Autoencoder Bestleistungen beim unüberwachten Training für visuelle Aufgaben, wie Objekterkennung, Detektion und semantische Segmentierung, erreicht werden können. In darauf aufbauenden Arbeiten [2, 3] wurde gezeigt, dass diese Form des unüberwachten Trainings sowohl für Vision Transformer (ViT) als auch für Convolutional Neural Networks (CNNs) nutzbar ist. Im Rahmen dieses Hauptseminars sollen Masked Autoencoder und deren Verwendung in ViT und CNNs aufbereitet werden.

Aufgabenstellung:

- Aufbereitung der Grundidee des Masked Autoencoder [1]
- Herausarbeiten der Unterschiede zur früheren Verwendung der Autoencoder zum Vortraining von Deep-Learning-Architekturen in der Anfangsphase des Deep Learning
- Aufarbeiten der Anwendung für ViT und CNNs [2, 3]
- Systematische Übersicht des State of the Art zu Ansätzen basierend auf [1]
- Abschätzung, wo sich diese Verfahren für die Robotik einsetzen lassen
- Vortrag im Rahmen des Hauptseminars

Geeignet für:

Bachelor- / Masterstudiengänge

Themengebiet / Schwerpunkte:

Deep Learning

Erforderliche Vorkenntnisse:

Guter Abschluss der Vorlesung „Neuroinformatik und Maschinelles Lernen“ und Erfahrungen im Bereich Deep Learning
oder erfolgreicher Abschluss der Vorlesung „Deep Learning for Computer Vision“

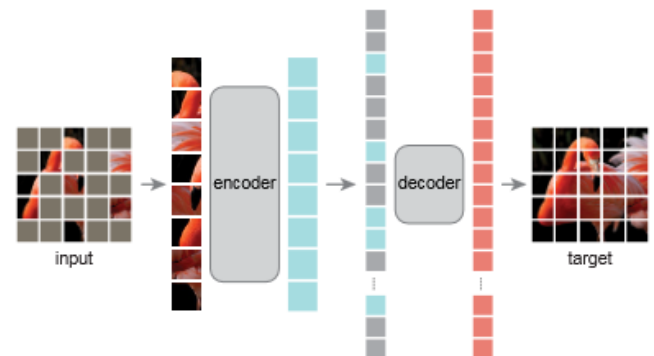
Zu verwendende Literatur:

- [1] He et al.: [Masked Autoencoders Are Scalable Vision Learners](#). CVPR, 2022.
 - [2] Dong et al.: [Bootstrapped Masked Autoencoders for Vision BERT Pretraining](#). ECCV, 2022.
 - [3] Li et al.: [Architecture-Agnostic Masked Image Modeling - From ViT back to CNN](#). arXiv, 2022.
- Elektronische Literaturdatenbank des FG NI&KR mit Recherchemöglichkeiten
 - Elektronische Konferenzproceedings-Datenbank des FG NI&KR
 - IEEE Recherchesystem www.ieeexplore.ieee.org (nur aus dem Uni-Netz bzw. via VPN)
 - Google Scholar scholar.google.com
 - Suche nach ähnlichen Publikationen connectedpapers.com, arxiv-sanity-lite.com
 - Proceedings der relevanten Konferenzen (NeurIPS, ICML, ICLR, IJCNN, WCCI, ICANN, CVPR, ICCV, ECCV, BMVC, AVSS, ICPR, ICIP, ...)

Betreuer: Dr. Markus Eisenbach (Markus.Eisenbach@tu-ilmeneau.de)

Betr. Hochschullehrer: Prof. Dr. H.-M. Groß

Bearbeiter: Ahmed Ramadan



Grundprinzip eines Masked Autoencoders. Bildquelle: [1]