



Elektrotechnisches Kolloquium

der Bergischen Universität Wuppertal

Die Fakultät für Elektrotechnik, Informationstechnik und Medientechnik lädt zur Teilnahme an folgender Vortragsveranstaltung mit anschließender Diskussion ein:

Beiträge zur Optimierung künstlicher neuronaler Netze

Es spricht

Kevin Kollek, M.Sc.

Lehrstuhl für Allgemeine Elektrotechnik und Theoretische Nachrichtentechnik

Prof. Dr. Anton Kummert

über das Thema

**Komprimierungsmethoden für künstliche neuronale Netze,
Realisierung von effizienten Algorithmen auf eingebetteten Systemen
und effiziente Approximation von Modelleleistungen**

Inhalt:

Der rasante technologische Fortschritt, insbesondere der Durchbruch im Bereich der Künstlichen Intelligenz, beeinflusst nahezu alle Wirtschaftssektoren – von autonomen Fahrzeugen über robotergestützte Fertigung bis hin zur medizinischen Diagnostik. Künstliche neuronale Netze ermöglichen dabei bemerkenswerte Innovationen, gehen jedoch mit erheblichen Herausforderungen einher. Vor allem der immense Rechen- und Energiebedarf, der sowohl beim Training als auch während der Inferenz entsteht, steht im Widerspruch zu den Zielen der Energiewende und einer ressourcenschonenden Entwicklung. Forschungsergebnisse zeigen jedoch großes Potenzial, die Effizienz neuronaler Netze zu steigern und deren Energieverbrauch deutlich zu reduzieren. Des Weiteren sind vor allem eingebettete Systeme häufig durch ihre begrenzten Ressourcen eingeschränkt. Dies erschwert die Echtzeitausführung komplexer neuronaler Netze und erfordert eine effiziente Anpassung der Algorithmen. Die Wahl der passenden Netzwerkarchitektur ist dabei besonders anspruchsvoll, da das Spektrum möglicher Operationen auf vielen eingebetteten Systemen stark limitiert ist.

In diesem Vortrag werden verschiedene effiziente und innovative Ansätze für die Optimierung von künstlichen neuronalen Netzen vorgestellt. Einerseits wird im ersten Teil gezeigt, wie ein leistungsstarkes neuronales Netz ausgewählt und komprimiert werden kann, wobei schrittweise verschiedene Bereiche des Netzes analysiert und sowohl redundante als auch irrelevante Informationen entfernt werden. Andererseits wird im zweiten Teil erläutert, wie neuronale Netze unter Ressourcen- und Echtzeitanforderungen auf einem eingebetteten Zielsystem implementiert werden, sodass eine automatisierte Verkehrszählung durchgeführt werden kann. Abschließend wird die Wahl der optimalen Architektur aus einem festgelegten Suchraum thematisiert. Hierzu werden neuartige Metriken zur Bewertung der Architekturqualität eingeführt und – gemeinsam mit etablierten Metriken – bereits in frühen Trainingsepochen ausgewertet, um so die leistungsstärksten Architekturen zu identifizieren.

Termin: 13.03.2025, 14 Uhr

Ort: Bergische Universität Wuppertal
Campus Freudenberg, Seminarraum FE 00.01