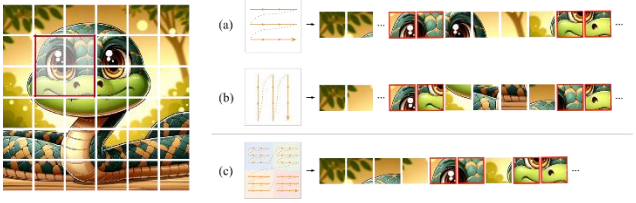
Hauptseminar

Thema: State Space Models für Bilderkennungsanwendungen

Neueste Fortschritte bei State Space Models, insbesondere Mamba [1], haben erhebliche Fortschritte bei der Modellierung langer Sequenzen in der Sprachverarbeitung erzielt und die Leistung der in diesem Bereich dominierenden Transformer-Architektur übertroffen. Erste Anwendungen im Bereich Computer Vision (ViM [2], VMamba [3]) konnten jedoch die Leistung traditioneller Convolutional Neural Networks (CNNs) und Vision Transformers (ViTs) nicht deutlich übertreffen. In [4] wurde gezeigt, dass der Schlüssel zur Verbesserung von Vision Mamba (ViM) in der Optimierung der Scanrichtungen für die Sequenzmodellierung liegt. Traditionelle ViM-Ansätze [2, 3] vernachlässigen die Erhaltung lokaler 2D-Abhängigkeiten (siehe Abbildung) und verlängern dadurch den Abstand zwischen benachbarten Token. Eine verbesserte lokale Scan-Strategie, die Bilder in verschiedene Fenster unterteilt und so lokale Abhängigkeiten effektiv erfasst, ohne die globale Perspektive zu vernachlässigen, behebt dieses Problem und führt zu deutlichen Verbesserungen auf dem ImageNet-Benchmarkdatensatz. Im Rahmen dieses Hauptseminars sollen das State Space Model Mamba [1, 5] und dessen Varianten im Vision-Bereich [2-4] aufbereitet werden.

Aufgabenstellung:

- Aufbereitung von Discrete State Space Models und Selective State Space Models [5]
- Aufarbeitung der Mamba-Architektur [1, 5]
- Aufzeigen der Vorteile gegenüber Rekurrenten Neuronalen Netzwerken (RNNs), Transformern und Convolutional Neural Networks (CNNs)
- Aufbereitung der geeigneten Anwendung der Mamba-Architektur im Vision-Bereich [2-4]
- Vortrag im Rahmen des Hauptseminars



Grundprinzip LocalMamba. Bildquelle: [4]

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] **Gu** et al.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv, **2023**.
[2] **Zhu** et al.: Vision Mamba: Efficient visual representation learning with bidirectional state space model. arXiv, **2024**.
[3] **Liu** et al.: VMamba: Visual state space model. arXiv, **2024**.
[4] **Huang** et al.: LocalMamba: Visual State Space Model with Windowed Selective Scan. arXiv, **2024**.
[5] **Grootendorst**: A Visual Guide to Mamba and State Space Models. Blogpost, **2024**.

- 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-ilmenau.de)
Betr. Hochschullehrer: Prof. Dr. H.-M. Groß
Bearbeiter: **offen**