

Hauptseminar

Thema: Erzeugung von Bildern durch Texteingaben

– Diffusion Models sind in der Lage realistische Bilder zu kreieren

Lange Zeit waren Generative Adversarial Networks (GANs) die bevorzugten Deep-Learning-Architekturen zur Erzeugung künstlicher Bilder, die von realen Bildern kaum zu unterscheiden sind. In [7] wurde jedoch gezeigt, dass Diffusion Models für diese Aufgabe sogar noch besser geeignet sind. Beeindruckende Resultate wurden durch zahlreiche Umsetzungen dieser Technik erzielt, wie z.B. durch Stable Diffusion [1] von Stability.ai, Imagen [2] von Google, und GLIDE [3] sowie DALL-E 2 [4] von OpenAI. Im Rahmen dieses Hauptseminars sollen Diffusion Models zur Bilderzeugung aufbereitet und in einer Implementierung demonstriert werden.

Aufgabenstellung:

- Aufbereiten der Grundidee der Diffusion Models [5-7]
- Aufbereitung des State of the Art zu Diffusion Models [8] mit Schwerpunkt auf bekannten Umsetzungen [1-4] zur Erzeugung von Bildern; Anwendungsmöglichkeiten aufzeigen
- Abschätzung, wo sich diese Verfahren für die Robotik einsetzen lassen
- Nutzung öffentlich verfügbarer Implementierungen zu Stable Diffusion [1] für eine prototypische Implementierung eines Demonstrators, der sich auf einem mobilen Roboter verwenden lässt
- Vortrag im Rahmen des Hauptseminars

Geeignet für:

Bachelor- / Masterstudiengänge

Themengebiet / Schwerpunkte:

Deep Learning



Durch Stable Diffusion erzeugte Bilder basierend auf (a) der Texteingabe „A painting of the last supper by Picasso.“ und (b) der Ausmaskierung der Person.
Bildquelle: [1]

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] Rombach et al.: [High-Resolution Image Synthesis with Latent Diffusion Models](#). CVPR, 2022.
- [2] Saharia et al.: [Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding](#). arXiv, 2022.
- [3] Nichol et al.: [GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models](#). arXiv, 2021.
- [4] Ramesh et al.: [Hierarchical Text-Conditional Image Generation with CLIP Latents](#). arXiv, 2022.
- [5] Sohl-Dickstein et al.: [Deep unsupervised learning using nonequilibrium thermodynamics](#). ICML, 2015.
- [6] Ho et al.: [Denosing diffusion probabilistic models](#). NeurIPS, 2020.
- [7] Dhariwal et al.: [Diffusion models beat GANs on image synthesis](#). NeurIPS, 2021.
- [8] Croitoru et al.: [Diffusion Models in Vision: A Survey](#). 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, ...)
- Zusammenfassungen auf Twitter zur Funktionsweise von Diffusion Models: [Tom Goldstein](#), [AI Pub](#)
- Öffentliche verfügbare Implementierungen zu Stable Diffusion: [Code](#), [optimierte Version](#), [Demo](#), [Web UI als JupyterNotebook in Colab](#), [API](#), [weitere Diffusion Models](#), [DearnStudio](#)

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

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

Bearbeiter: Eline Laura Elissa Vervoort