

KI Entwicklungen April 2023

Im Bereich der Künstlichen Intelligenz (KI) gab es im April 2023 spannende Fortschritte, die auch für die Film- und Medienproduktion von Bedeutung sind. Hier sind die wichtigsten Entwicklungen aus den Bereichen Text-, Bild- und Video-Generierung sowie 3D und NeRF zusammengefasst:

Chatbots:

- Open-Source-Chatbot [Vicuna](#): basiert auf dem LLaMA-Modell und wurde mit Hilfe von ShareGPT-gesammelten Unterhaltungen feinetuned. Die Ergebnisse der Vorabauswertung sind erstaunlich, da Vicuna-13B etwa 90% der Qualität von ChatGPT erreicht, wodurch es Modelle wie LLaMA und Alpaca übertreffen kann. Die öffentliche Verfügbarkeit des Codes und der Gewichte sowie eine Online-Demo machen Vicuna für interessierte Entwickler und Forscher zugänglich.
- Darüber hinaus hat das Unternehmen Stability AI, welches an Stable Diffusion beteiligt war, eine eigene Version des Sprachmodells [StableLM](#) veröffentlicht, welches ebenfalls auf LLaMA und Vicuna basiert.

Video:

- [Gen-2-Modell](#) in RunwayML: ist in der Lage, Videos in verschiedenen Styles zu synthetisieren und zu erstellen. Das Modell beherrscht acht verschiedene Modi, einschließlich Text-to-Video, Image-to-Video und Stylization.
- Open-Source-Modell [VideoCrafter](#) auf Hugging Face: enthält drei verschiedene Arten von Modellen, einschließlich eines Text-to-Video-Modells, das auf latenten Video-Diffusionsmodellen basiert, einem personalisierten Text-to-Video Modell und einem Video-Generation-Modell mit mehr Condition Controls.
- Modell von Nvidia [VideoLDM](#): ermöglicht Text-to-Video-Generierung und auch die Veränderung eines Driving Videos. Die Ergebnisse sind aufgrund der zeitlich sehr konstanten und optisch beeindruckenden Ergebnisse sehr bemerkenswert.

3D / NeRF:

- Paper von Nvidia "[GeNVS](#): Generative Novel View Synthesis with 3D-Aware Diffusion Models": beschreibt eine neue Methode zur generativen Synthese von Ansichten aus 3D-Modellen. Die Methode nutzt bestehende 2D-Diffusionsmodelle, um aus wenigen Eingabebildern eine Menge von Renderings zu erzeugen, die plausibel und divers sind. Die Methode integriert jedoch auch Geometriprioritäten in Form eines 3D-Feature-Volumens, was die Fähigkeit des Modells zur Generierung von konsistenten Ansichten verbessert. Die Autoren demonstrieren die Effektivität ihrer Methode anhand von synthetischen Renderings und realen Szenen.
- Paper von Google "[Zip-NeRF](#): Anti-Aliased Grid-Based Neural Radiance Fields": befasst sich mit der Beschleunigung von NeRF-Training durch die Verwendung von gitterbasierten Repräsentationen. Die Autoren zeigen, wie sich durch Kombination von gitterbasierten Techniken mit einer anderen Methode namens "mip-NeRF 360" die Fehlerrate reduzieren und die Trainingszeit verkürzen lässt.
- Forscher der UC Berkeley haben eine Methode namens "[Instruct-NeRF2NeRF](#)" entwickelt, mit der sich NeRFs per Texteingabe bearbeiten lassen. Die Methode erlaubt es beispielsweise, Gesichter zu bearbeiten, Tageszeiten zu ändern oder die Kleidung von Personen zu wechseln. Allerdings erfordert die Methode eine große Menge an Videospeicher.

Segmentierung von Bildern:

- AI-Modell von Meta AI "[Segment Anything](#) (SAM)": ist in der Lage, jedes Objekt in einem Bild mit einem einzigen Klick auszuschneiden. Das System kann ohne zusätzliches Training auf unbekannte Objekte und Bilder angewendet werden und bietet damit ein hohes Maß an Generalisierung.

Insgesamt sind diese Entwicklungen bedeutende Fortschritte im Bereich der KI und könnten Anwendungen in der Text-Generierung, 3D-Modellierung, der Computergrafik und der Bildbearbeitung ermöglichen.