Projektdokumentation

Vergleich cloudbasierter und lokaler KI-Videogenerierung am Beispiel von Sora und WAN 2.2

1. Einleitung

Die schnelle Entwicklung generativer KI-Modelle hat in den letzten Jahren zu erheblichen Fortschritten im Bereich der Videogenerierung geführt. Neben cloudbasierten Systemen großer Anbieter entstehen immer mehr lokal ausführbare Open-Source-Modelle, die viele Möglichkeiten, aber auch Einschränkungen mit sich bringen.

Ziel des Projekts ist ein Vergleich zwischen einem lokal betriebenen KI-Videomodell (WAN 2.2 in ComfyUI) und einem cloudbasierten Video-Foundation-Model (Sora). Der Fokus liegt dabei nicht auf visueller Ästhetik allein, sondern auf zeitlicher Kohärenz, Identitätsstabilität, physikalischer Plausibilität und Steuerbarkeit.

2. Zielsetzung

Die Dokumentation verfolgt folgende Ziele:

- Analyse der grundlegenden technischen Unterschiede zwischen lokalem und cloudbasiertem Ansatz
- Bewertung der Stärken und Schwächen beider Systeme anhand identischer Textprompts
- Untersuchung der Grenzen lokaler Videogenerierung in Bezug auf Zeit, Auflösung und Handlung
- Ableitung von praxisnahen Einsatzmöglichkeiten für beide Modelle

3. Systemübersicht

3.1 WAN 2.2 (ComfyUI lokal)

WAN 2.2 ist ein diffusionsbasiertes KI-Videomodell, das lokal auf einer GPU ausgeführt wird. Die Integration in ComfyUI erlaubt eine einfache und visualisierte Steuerung über Nodes, darunter Prompting, Seed-Kontrolle, Framelänge, Auflösung und Nachbereitung.

Charakteristika:

- Lokale Ausführung
- Hohe Kontrolle und Reproduzierbarkeit
- Begrenzte Videolänge
- Starke Abhängigkeit von VRAM, FPS, und Auflösung

3.2 Sora (cloudbasiert)

Sora ist ein sehr groß skaliertes, cloudbasiertes KI-Videomodell, das Video nicht als isolierte Framefolge, sondern als zusammenhängende zeitlich-räumliche Struktur generiert (Diffusion Transformer). Der Nutzer interagiert ausschließlich über Textprompts, ohne direkten Zugriff auf technische Parameter.

Charakteristika:

- Cloudbasierte Ausführung
- Hohes Maß an zeitlicher Kohärenz
- Stabile Identitäten über lange Sequenzen
- Geringe direkte Kontrolle über Details

4. Durchführung

4.1 Prompt-Design

Um einen guten Vergleich zu ermöglichen, wurden modellneutrale Textprompts verwendet:

Beispiele:

- Einfache Objektinteraktionen
- Kontinuierliche menschliche Bewegung
- Mehrschrittige Handlungsfolgen

4.2 Technische Parameter (WAN 2.2)

- Auflösung: 512 x 512 (Standard)
- Bildrate: 16 & 24 fps
- Länge: 3-5 Sekunde
- Noise_Seed: 1071099550038382

Sora wurde mit denselben Prompts und vergleichbarer Szenenlänge getestet (Im Zeitraum vom 15-17.12.2025).

4.3 Wan2.2 lokal auf einem Macbook

WAN 2.2 wurde ausschließlich für Linux + NVIDIA-GPUs (CUDA) gebaut. MacOS (CPU/MPS) ist kein unterstütztes Ziel, weder offiziell noch inoffiziell.

Folgende Probleme sind beim Versuch WAN2.2 auf lokal auf dem Macbook zum Laufen zu bringen:

- CUDA ist fest im Code verdrahtet
- Die VAE ist CUDA-only entworfen
- Das Modell ist extrem groß

5. Bewertungskriterien

Die Ergebnisse wurden anhand folgender Kriterien bewertet:

- Identität über Zeit (Stabilität von Personen/Objekten)
- Zeitliche Kohärenz (flüssige Bewegungen, keine Sprünge)
- Physik & Kausalität (logische Abfolge von Aktionen)
- Kamera- und Raumkonsistenz
- Prompt-Treue

6. Ergebnisse

6.1 WAN 2.2

WAN 2.2 erzeugt visuell ansprechende Kurzvideos mit hoher stilistischer Kontrolle. Besonders stark ist das Modell bei atmosphärischen Szenen, abstrakten Motiven und kurzen Bewegungen. Deutliche Schwächen zeigen sich jedoch bei längeren Sequenzen, komplexen Handlungen und der Stabilität von Identitäten über mehrere Sekunden hinweg.

6.2 Sora

Sora zeigt eine deutlich höhere zeitliche und semantische Kohärenz. Mehrschrittige Handlungen werden in logischer Reihenfolge ausgeführt, Identitäten bleiben stabil, und physikalische Zusammenhänge sind meist plausibel. Demgegenüber stehen eine geringere Steuerbarkeit sowie fehlende Reproduzierbarkeit einzelner Ergebnisse

7. Diskussion

Der Vergleich zeigt, dass beide Systeme unterschiedliche Problemräume andressieren. Während WAN 2.2 als lokales Werkzeug vor allem für Experimente, Forschung und kontrollierte Generierung geeignet ist, stellt Sora einen Ansatz dar, der Videogenerierung als Weltmodell mit Zeitverständnis interpretiert.

8. Fazit

- Wan 2.2 und Sora sind nicht direkte Konkurrenten, sondern repräsentieren zwei Extreme im aktuellen KI-Video-Ökosystem. WAN 2.2 bietet Kontrolle, Transparenz und lokale Ausführbarkeit, während Sora durch Kohärenz, Narration und zeitliche Stabilität überzeugt.
- Für kurze, kontrollierte Clips ist WAN 2.2 ein leistungsfähiges Werkzeug. Für längere, zusammenhängende Videos mit komplexer Handlung zeigt sich Sora derzeit klar überlegen.

9. Gelerntes und Erkenntnisse

Im Verlauf des Projekts wurden technische und auch grundlegende Erkenntnisse gewonnen.

9.1 Technische Erkenntnisse

Eine zentrale Erkenntnis war, dass die Qualität von KI-generierten Videos nicht nur durch Auflösung und visueller Schärfe bestimmt wird, sondern auch durch zeitliche Kohärenz und Identitätsstabilität. Gerade bei lokalen Modellen wie WAN2.2 sieht man, das kurze Clips mit einem unkomplizierten Prompt und niedrigerer Auflösung zu besseren Ergebnissen führen.

Darüber hinaus wurde klar, wie sehr Frameanzahl, Bildrate und VRAM zusammenhängen. Kleine Änderungen in FPS oder Auflösung konnten erhebliche Auswirkungen auf die Stabilität und Rechenaufwand haben.

Zu Beginn des Projektes wurde auch klar, dass das Betreiben von generativen KI-Modellen auf Mac-Books nahezu unmöglich ist.

9.2 Gelerntes

Das Projekt hat das Verständnis dafür geschärft, wie sich aktuelle KI-Videogenerationsmodelle anfühlen, wenn man mit ihnen arbeitet. Viele theoretische Begriffe wie Identität über Zeit oder Kausalität wurden erst durch die praktische Erfahrung wirklich klar.

Die praktische Arbeit hat auch meine kritische Einschätzung von KI generierten Videos verbessert, aber auch die Wertschätzung, sehr guter KI generierter Videos gesteigert.

10. Sonstiges

- Offizielle ComfyUI Dokumentation für die Portable Installations-Variante: https://docs.comfy.org/installation/comfyui_portable_windows?utm_source=chatgpt.com
- Dokumentation für Desktop-Installer-Version (für das Projekt verwendet): https://docs.comfy.org/installation/desktop/windows?utm_source=chatgpt.com