Künstliche Intelligenz für Video und Audio Technologien

Digitale AV Technik, MIB 5

Übersicht

- GAN
- Stable Diffusion

Generative Adversarial Networks (GANs)

- Ein maschinelles Lernmodell mit zwei konkurrierenden Netzwerken:
  - **Generator**
  - **Diskriminator**
- Entwickelt von Ian Goodfellow et al. (NeurIPS2014)
- Anwendung: Bildgenerierung, Stiltransfer, Datenaugmentation, u.v.m.

Grundidee: Generator vs. Diskriminator

Zwei Netzwerke im Wettstreit:

1. **Generator**
    - Ziel: Realistisch aussehende Daten erzeugen.
    - Erzeugt zufällige Daten zu Beginn.
    - Lernt, echte Daten zu imitieren.
2. **Diskriminator**
    - Ziel: Zwischen echten und generierten Daten unterscheiden.
    - Prüft die Arbeit des Generators.
    - Lernt, echte Daten von generierten zu unterscheiden.

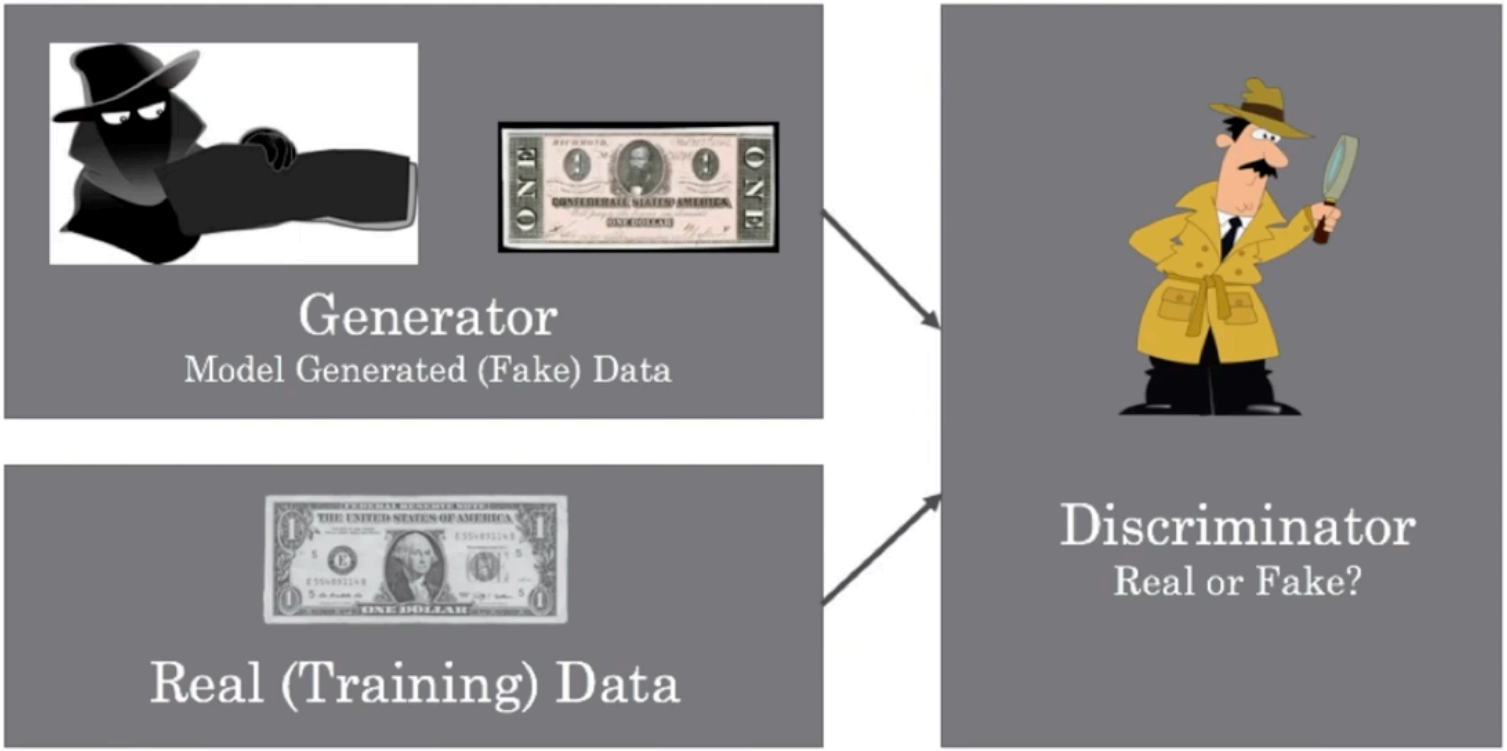
Spiel zwischen Generator und Diskriminator

- **Generator:** Täuscht den Diskriminator mit besseren Fälschungen.
- **Diskriminator:** Lernt, die Täuschungen zu erkennen.

Ergebnis: Beide Netzwerke verbessern sich stetig.

> Ziel: Der Generator erzeugt Daten, die ununterscheidbar von echten Daten sind.

Beispiel mit Illustration



> Quelle: Gentle Intro to Generative Adversarial Networks - Part 1 (GANs) by WelcomeAIOverlords (YouTube)

Analogie: Künstler vs. Kritiker 🎨🧐

- Der **Generator** ist ein Künstler:
  - Versucht, Kunstwerke zu erstellen, die wie echte aussehen.
- Der **Diskriminator** ist ein Kunstkritiker:
  - Erkennt echte Kunstwerke und entlarvt Fälschungen.

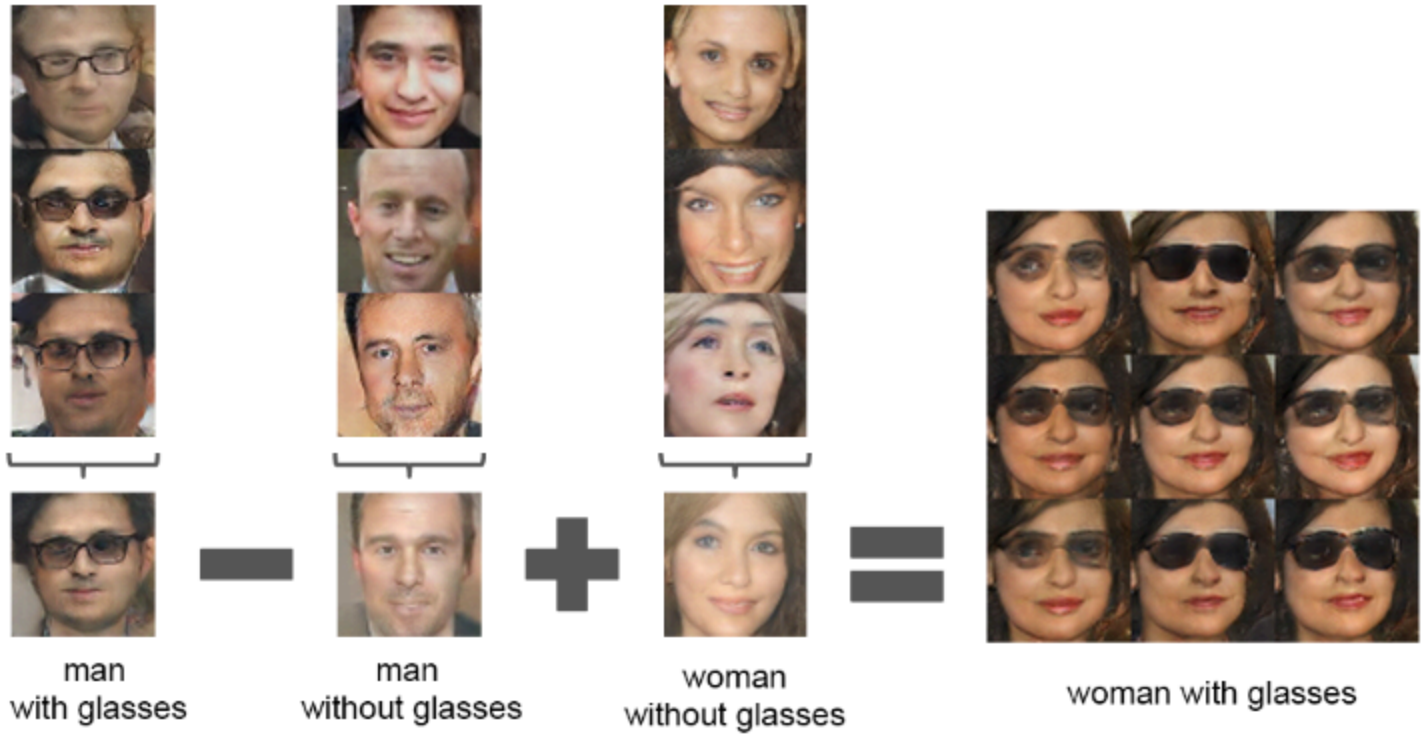
Training:

- Der Künstler wird immer besser, bis seine Werke nicht mehr von echten zu unterscheiden sind.

Ende des Trainings

- Der Diskriminator kann nur noch zufällig zwischen echten und generierten Daten unterscheiden (50 % Genauigkeit).
- Die vom Generator erzeugten Daten sind nahezu perfekt.
- Es kann sehr lange dauern und es hängt stark von den Trainingsdaten ab.

Arithmetik



- Quelle: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, *Alec Radford, Luke Metz, Soumith Chintala*, ICLR 2016 pdf

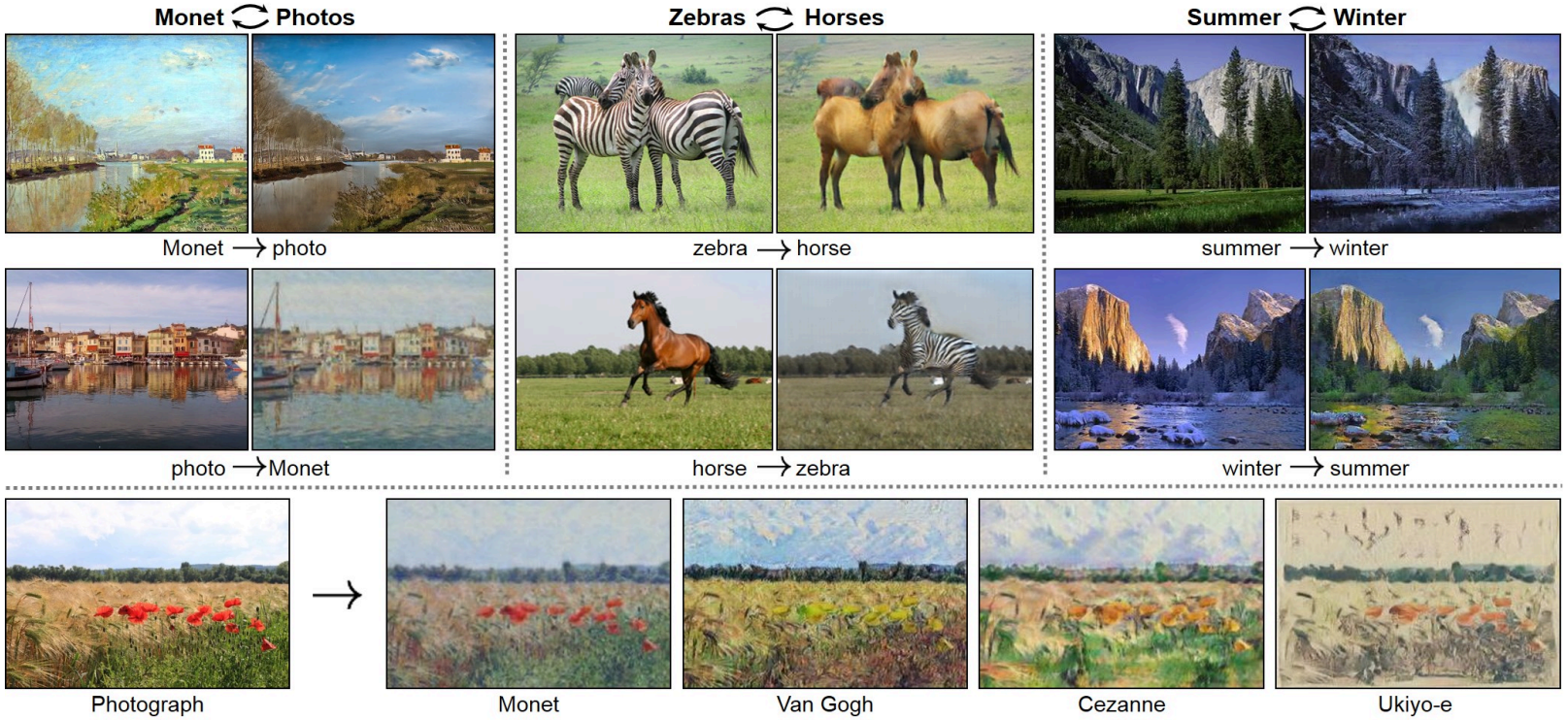
Varianten

- StyleGAN, StyleGAN2 Demo
- CycleGAN
- 3D GANs
- Audio GAN

StyleGAN

A Style-Based Generator Architecture for Generative Adversarial Networks, *Tero Karras (NVIDIA), Samuli Laine (NVIDIA), Timo Aila (NVIDIA)*, CVPR 2019

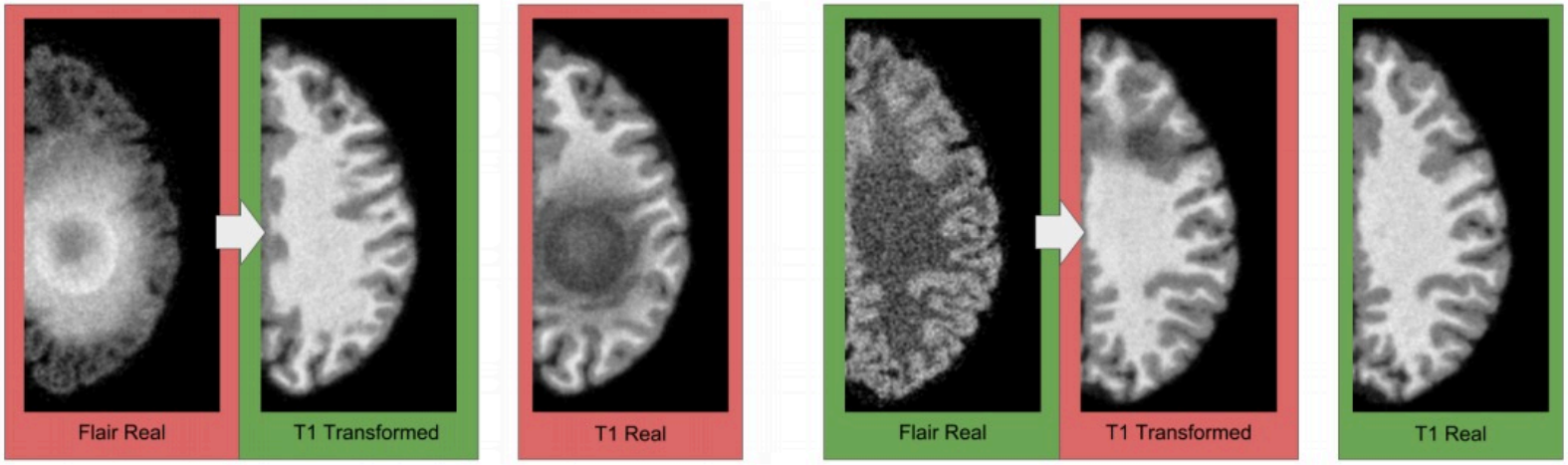
CycleGAN



Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,

Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, ICCV 2017

Ein Hinweis zu CycleGAN



(a) A translation removing tumors (b) A translation adding tumors

> "CycleGAN should only be used with great care and calibration in domains where critical decisions are to be taken based on its output."

Quelle: CycleGAN Projektseite

3D GANs

- Idee: Bilder aus mehreren Perspektiven des selben Objekts erzeugen und dann 3D Rekonstruktion anwenden
- Stand der Technik 2021 mit Video
- Nachteil: Training extrem aufwendig

3D GANs (Voxel-basiert)

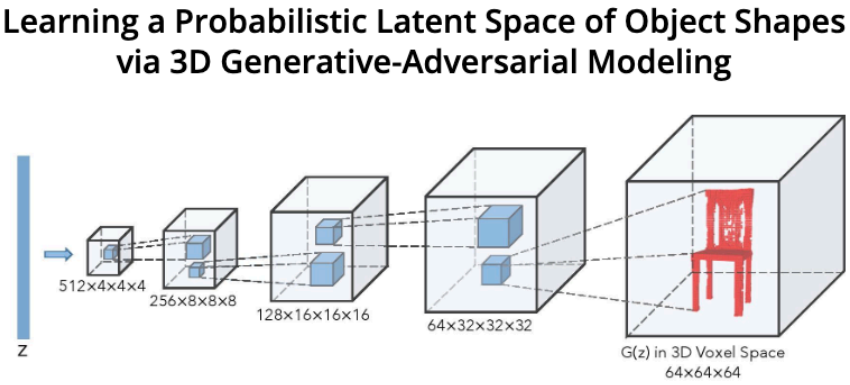
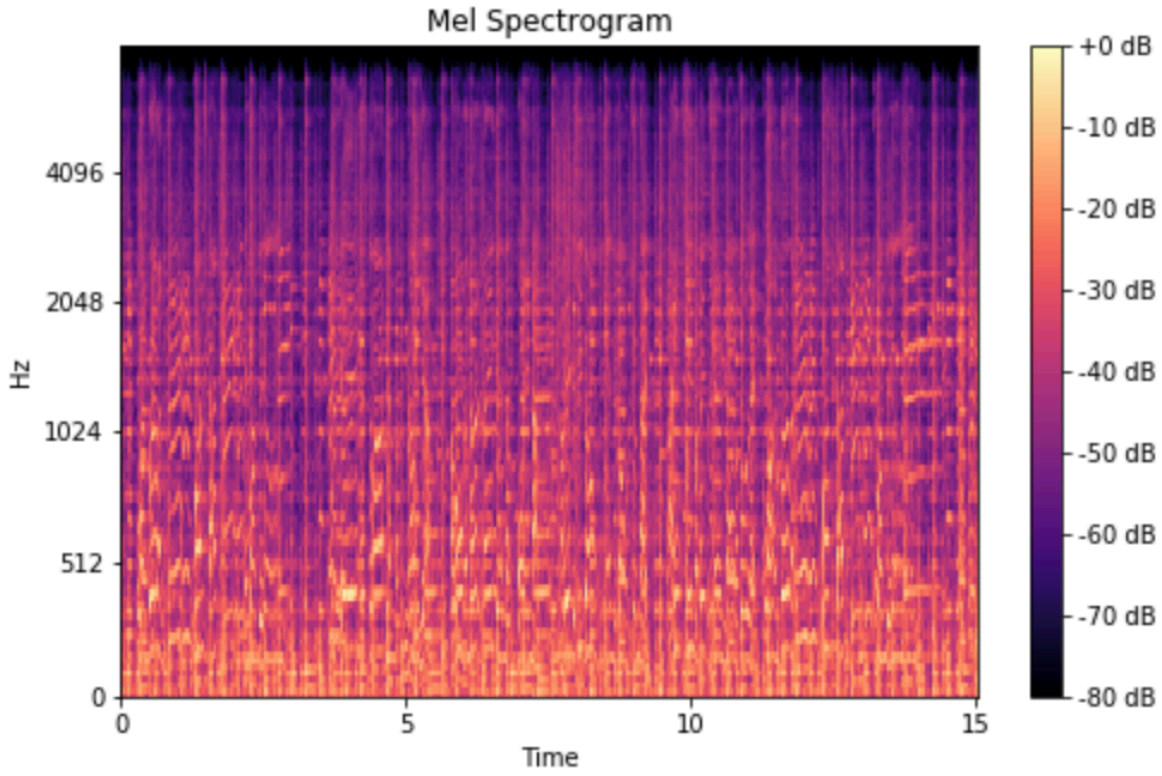


Figure 1: The generator of 3D Generative Adversarial Networks (3D-GAN)

Figure 2: Shapes synthesized by 3D-GAN

Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, *Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum*, NeurIPS 2016

Audio GANs (früherer Ansatz)



Quelle: Understanding the Mel Spectrogram by Leland Roberts

AudioGANs (Beispiele)

Stand der Technik 2021 mit Hörbeispielen:

Unofficial Parallel WaveGAN implementation demo

GANs vs. Stable Diffusion

- **Beide sind generative Modelle**, aber mit unterschiedlichen Ansätzen.
- Ziel: Realistische Daten, wie Bilder oder Texte, erzeugen.
- Unterschiedliche Prinzipien:
  - GANs: Wettbewerb zwischen Generator und Diskriminator.
  - Stable Diffusion: Iterative Umkehr eines Rauschprozesses.

Stable Diffusion: Ein neuer Ansatz

- **Diffusionsmodell**:
  - Bilder werden **schrittweise aus Rauschen rekonstruiert**.
  - Training: Das Modell lernt, den **Rauschprozess umzukehren**.
  - Erfunden in München: **High-Resolution Image Synthesis with Latent Diffusion Models**, *Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer*, CVPR 2022

Stable Diffusion im Detail

https://jalammar.github.io/illustrated-stable-diffusion/

Jeder liest für sich den Artikel und sammelt Verständnisfragen

Vorteile von Stable Diffusion

- **Stabilität:** Robusteres Training im Vergleich zu GANs.
- **Effizienz:** Reduzierte Ressourcenanforderungen durch optimierte Architektur.
- **Kontrollierbarkeit:** Ermöglicht gezielte Generierung (z. B. Text-zu-Bild).

Stable Diffusion (v1.5 from 2022)

Prompt: "frog with melon hat"

Stable Diffusion 1.5 at huggingface

Stable Diffusion (v3.5 from 2024)

Prompt: "frog with melon hat"

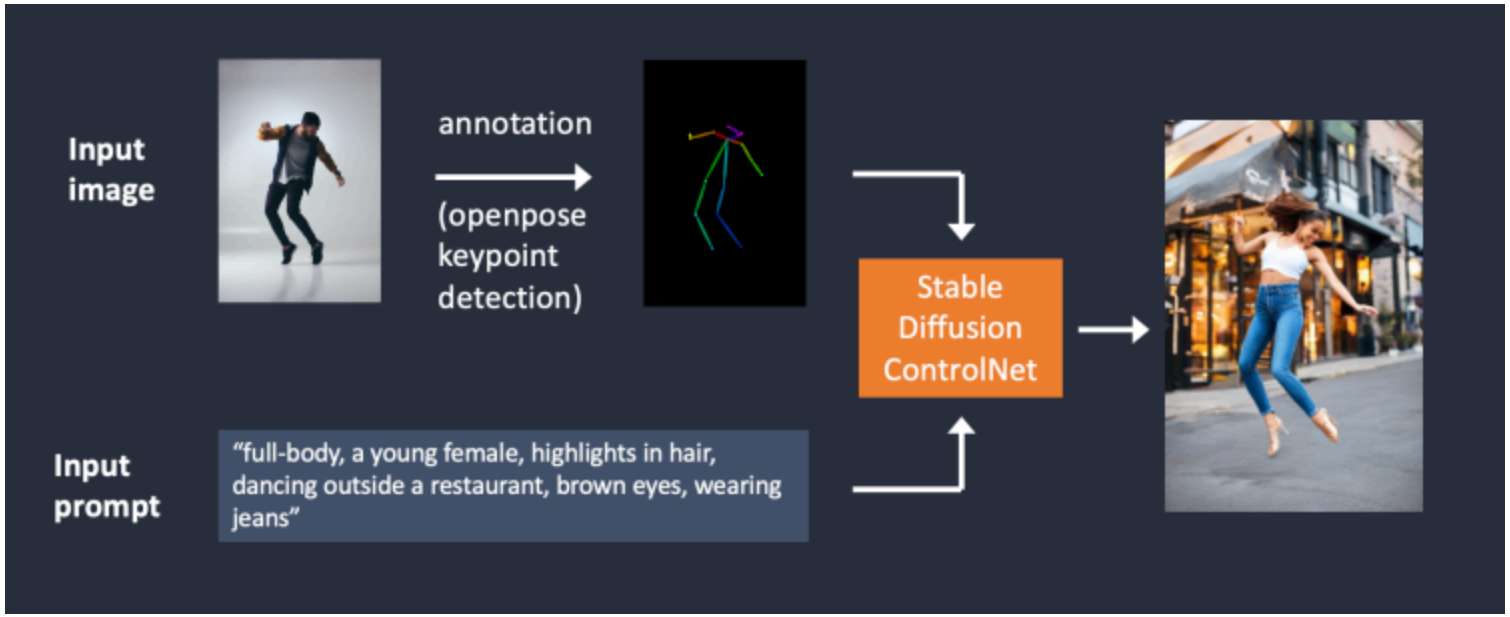
StabilityAI Stable Diffusion 3.5 large

FLUX.1 [dev] (2025)

Prompt: "frog with melon hat"

Black Forest Labs

Stable Diffusion steuern (ControlNet)

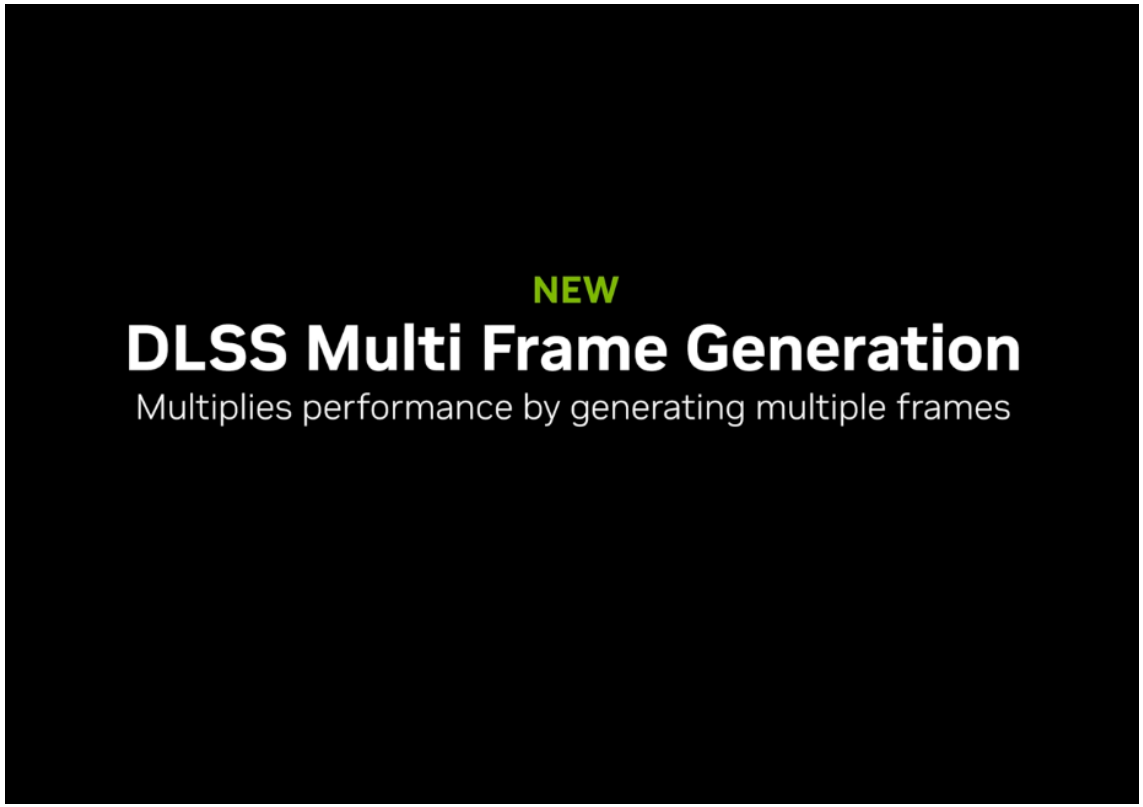


Bildquelle: ControlNet: A Complete Guide, Updated July 7, 2024, By Andrew

Text zu 3D mit StableDiffusion + NeRF

- Stand der Technik 2021 mit Video
- Stand der Technik 2024: Sora von OpenAI, veo 2 von Google

The future: Frame generation



NVIDIA DLSS4, video