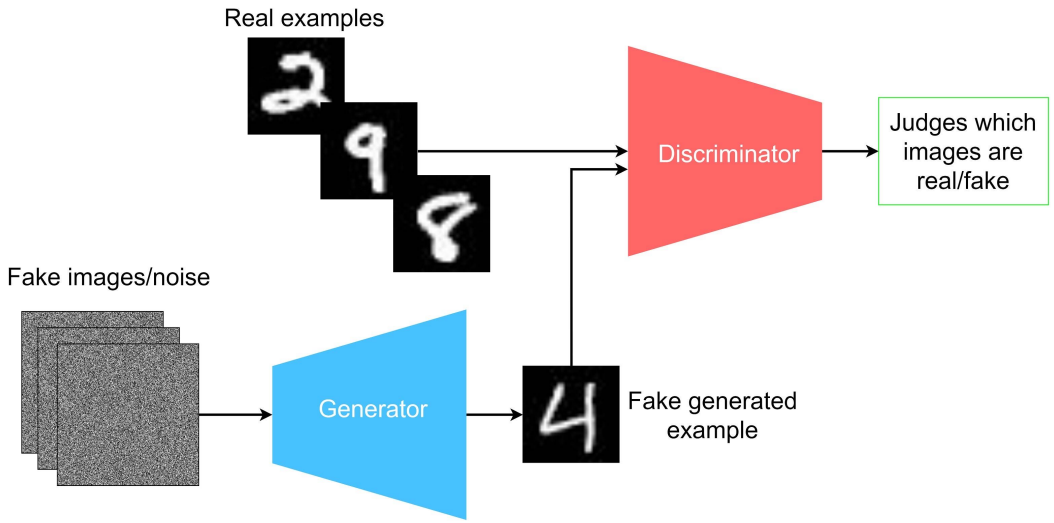
Image Generation with Foundation Models

Entwicklung der Image Generation

- 1960 → AARON
  - Erstes autonomes Kunstprogramm – regelbasiert, kein maschinelles Lernen.
- 2010 → GANs (Generative Adversarial Networks)
  - Zwei konkurrierende Netze erzeugen erstmals realistische Bilder.
- 2021 → DALL·E (OpenAI)
  - Text-zu-Bild-Modell mit Transformer-Architektur.
- 2022 → DALL·E 2
  - Wechsel zu Diffusionsmodellen – höhere Auflösung, bessere Kohärenz.
- 2023 → DALL·E 3
  - Integriert GPT -4 für präzisere Text-Bild-Übersetzung und Kontextverständnis.

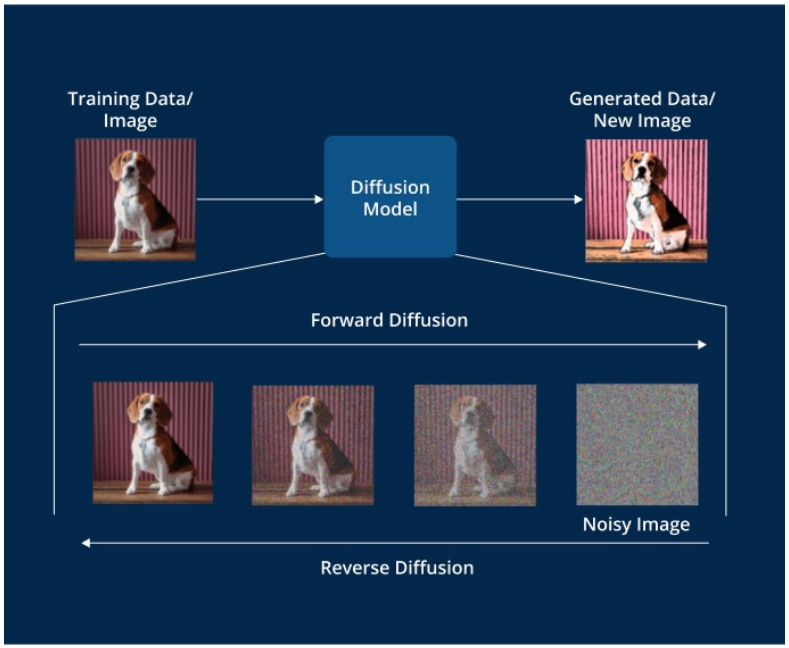
GANs – Grundprinzip

- Besteht aus zwei Neuralen Netzwerken
  - Der Generator
  - Der Diskriminator
- Erzeugen Daten durch Konkurrenz
- Schwer zu trainieren, neigen zu Instabilität



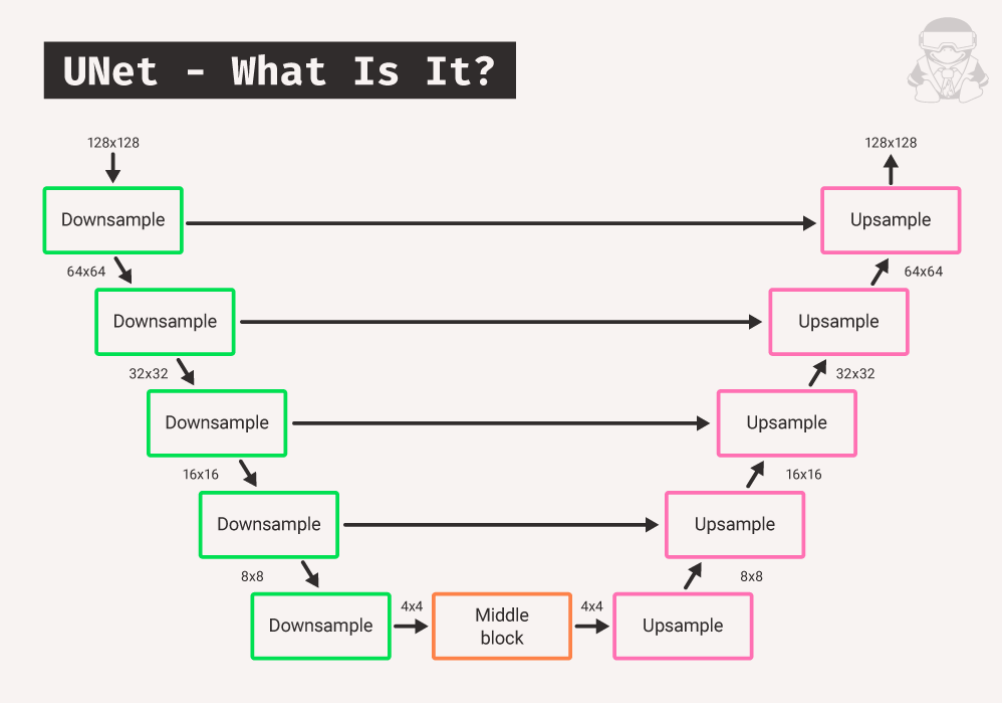
Diffusionsmodel – Grundprinzip

- CLIP wandelt den Text in einen Vektor um.
- Diffusion zerstört ein Bild mit Gauß-Rauschen → lernt, es wieder zusammenzusetzen.
- Beim Sampling nutzt das Netz CLIPs Bedeutung, um aus Rauschen ein sinnvolles Bild zu „rekonstruieren“.
- VAE macht daraus ein richtiges Bild.



U-Net

- Architektur, um Bilder zu verstehen und wieder aufzubauen
- Encoder
  - Bild wird immer verkleinert
  - Wichtigsten Strukturen werden beibehalten
- Decoder
  - Netz baut das Bild wieder auf
  - Sauberer, detaillierter, oder „Entrauscht“
- Skip Connection
  - Sorgen dafür das feine Details nicht aus früheren Schichten verloren gehen



Diffusionsmodel – Ablauf

1. Textverarbeitung
   - Text wird umgewandelt in Vektor
2. Startzustand
   - Zufälliges Latente Bild voller Rauschen wird generiert
3. Diffusion (U-Net arbeitet)
   - **U-Net Encoder** liest das verrauschte latente Bild, reduziert es schrittweise (Downsampling) und sammelt Kontextinformationen.
   - **Skip Connections** leiten Details aus frühen Schichten an spätere weiter
   - **U-Net Decoder** baut das Bild wieder auf (Upsampling) und versucht, das Rauschen zu entfernen.
4. Decoder
   - Das VAE-Decoder-Netz Upsampled es stufenweise bis ein fertiges Bild zu sehen ist

Flow-Matching Model – Grundprinzip

- Kontinuierliche Transformation zwischen Rauschen und Daten (Ein Fluss)
- Keine diskreten Diffusionsschritte nötig → effizientere Generierung
- Funktion wie sich Datenpunkte im Laufe der Zeit Bewegen
- stabileren und schnelleren Samplingprozessen
- Deterministisch statt stochastisch

Diffusion Model VS Flow-Matching Model

Aspekt	Diffusion	Flow
Mathematische Grundlage	Stochastischer Prozess (SDE)	Deterministischer Fluss (ODE)
Generierung	Schrittweise Entrauschung	Direkte kontinuierliche Transformation
Training	Modelliert Rauschentfernung	Modelliert Trajektorien zwischen Rauschen und Daten
Geschwindigkeit	Langsamer (viele Schritte)	Schneller (oft weniger Integration schritte)
Stabilität	Sehr stabil, aber rechen intensiv	Potenziell schneller, aber etwas sensibler
Beispiele	Stable Diffusion, Imagen	Flow Matching Rectified Flow, Flow Matching Diffusion Models

Diffusion Model VS Flow-Matching Model

Kriterium	Diffusion	Flow
Trainingsaufwand	Hoch (viele Schritte, robust)	Mittel (weniger Schritte, sensibel)
Energieverbrauch (Training)	Sehr Hoch	Mittel
Energieverbrauch (Sampling)	Hoch (viele Iterationen)	Niedrig (direkte Integration)
Rechengenauigkeit	Tolerant gegenüber Rauschen	Präzise, empfindlich bei Fehlern

Stable Diffusion Beispiel

Fazit

- Diffusionsmodelle sind heute der Standard in der Bildgenerierung
- Bieten hohe Qualität, Stabilität und Kontrolle bei der Bildproduktion
- Flow-Matching-Modelle: neue Ansätze für effizientere und deterministische Generierung
- Fokus auf Nachhaltigkeit, Schnelligkeit und Generalisation
- Stable Diffusion: Beispiel für offene, skalierbare und gemeinschaftsbasierte KI-Innovation
- Zukunft: bessere Effizienz, Zugänglichkeit und verantwortungsvolle Nutzung von KI-Modellen

Quellen

- https://www.stefanjacob.de/geschichte/die-historische-entwicklung-von-ai-images-generatoren/
- https://developer.ibm.com/articles/generative-adversarial-networks-explained/
- https://www.leewayhertz.com/diffusion-models/
- https://eviltux.com/2024/08/11/training-a-u-net-model-from-scratch/
- https://huggingface.co/stabilityai/stable-diffusion-2-base
- https://github.com/AUTOMATIC1111/stable-diffusion-webui
- https://www.youtube.com/watch?v=ZXiruGOCn9s
- https://www.youtube.com/watch?v=wnuWqG18FVU
- https://www.youtube.com/watch?v=zc5NTeJbk-k
- https://www.youtube.com/watch?v=firXjwZ_6KI&t=810s
- https://www.youtube.com/watch?v=1pgiu--4W3I&t=82s
- https://www.youtube.com/watch?v=KcSXcpluDe4