

KI-Videos mit Stable Video Diffusion erstellen



Auf Pollux Labs konntest du bereits viel über Bildgenerierung mit ChatGPT und deepAI erfahren – aber wie sieht es mit Videos aus? Stable Diffusion hat im November 2023 eine [frühe Version einer Software](#) veröffentlicht, die auf Basis eines Ausgangsbilds ein kleines KI-Video erstellen kann. **In diesem Tutorial erfährst du, wie das geht.**

Ein kurzer Hinweis: Dieses Tutorial behandelt eine frühe Version und ist auf dem Stand Ende November '23. Wie du weißt, entwickelt sich die KI-Welt rasant – also möglicherweise hat sie sich schon weitergedreht, wenn du diese Zeilen hier liest.

Das kann Stable Video Diffusion

Gegenwärtig kannst du mit Stable Video Diffusion eine kurze Video-Sequenz erstellen, die entweder 14 oder 25 Frames lang ist. Als Ausgangspunkt dient ein einzelnes Bild, das zwingend im Format 576×1024 Pixel vorliegen muss. Künftig ist auch eine Version geplant, die auf Basis eines Textprompts Videos erstellen kann, für die du dich bereits in eine Warteliste eintragen kannst.

Testweise habe ich eine kleine Animation einer Gummiente auf einem Fluss erstellt. Hier das Ergebnis:

Das Ausgangsbild hierfür habe ich mit ChatGPT erstellt (Hier ein [Tutorial, wie du per API Bilder mit ChatGPT erzeugen](#) kannst).

So erstellst du ein KI-Video auf Basis eines Bilds

Du findest auf Github das [Repository von Stable Video Diffusion](#). Wenn du Erfahrung mit dem Klonen und Einrichten von Repositories hast, kannst du lokal auf deinem Computer installieren. In diesem [Video auf Youtube](#) erhältst du weitere Information zum Download des benötigten KI-Modells und zu ComfyUI – einer Oberfläche, mit der du Stable Diffusion verwenden kannst.

Für einen ersten Test gibt es jedoch eine einfachere Möglichkeit, die dir viel Zeit spart.

Auf der Webseite von Replicate findest du eine [Online-Version von Stable Video Diffusion](#), mit der du ohne großen Aufwand experimentieren kannst. Du musst lediglich ein Ausgangsbild im erwähnten Format von 576×1024 Pixeln hochladen und für einen ersten Test auf **Run** klicken. Das Video wird dann direkt in deinem Browser erstellt:

Input

Form Node.js Python Elixir HTTP Cog Docker



input_image* File

Drop a file or click to upload

DALL·E 2023-11-27 10:49:07 - A vibrant yellow rubber duck floating serenely on a river. The rubber duck is classic in design, with a...

Take a photo with your webcam

Reset

Run

Output

Running...

Cancel

Hide logs

```
Set consts
Loaded model
##### Sampling setting #####
Sampler: EulerEDMSampler
Discretization: EPMDiscretization
Guider: LinearPredictionGuider
Sampling with EulerEDMSampler for 26 steps: 0% | 0/26 [00:00:07, 71t/s]
Sampling with EulerEDMSampler for 26 steps: 4% | 1/26 [00:01:00:42, 1.70s/it]
Sampling with EulerEDMSampler for 26 steps: 8% | 2/26 [00:03:00:42, 1.76s/it]
Sampling with EulerEDMSampler for 26 steps: 12% | 3/26 [00:05:00:41, 1.78s/it]
Sampling with EulerEDMSampler for 26 steps: 15% | 4/26 [00:07:00:39, 1.80s/it]
```

Auf der rechten Seite erscheint dann das fertige Video (oder eher eine kurze Animation), die auf Basis des Bilds links erzeugt wurde. Hinweis: Ein von ChatGPT erzeugtes Bild hat im Landscape-Format nicht ganz die richtige Größe für die Video-Generierung. Passe die Auflösung deshalb vor dem Upload manuell an.

Wie du im Video oben siehst, versucht Stable Video Diffusion zu erkennen, was sich auf dem statischen Bild für eine Animation eignen könnte. In unserem Fall ist das natürlich das Wasser des Flusses und die Ente, die darauf schwimmt. **Selbst kannst du noch keine Vorgaben machen, was im Video passieren soll** – also ob z.B. ein Stück Treibholz ins Bild schwimmt.

Voreingestellt sind 14 Frames, die mit den voreingestellten Frames pro Sekunde ein KI-Video von circa 2 Sekunden Länge ergeben. Du kannst aber die Videolänge auch auf 25 Frames erhöhen. Außerdem kannst du noch die Frames per Second definieren und auch mit weiteren Werten, wie der „Menge“ an Bewegung, die im fertigen Video zu sehen sein soll,

experimentieren.

Hier das Ergebnis eines **Baums im Sturm** mit 25 Frames, 15 Frames pro Sekunde und einer **motion_bucket_id** von 255:

Wie du hier siehst, geht es zwar rund mit den Wolken – der Baum scheint vom Sturm allerdings eher unbeeindruckt zu sein, abgesehen von ein paar herumwehenden Blättern. Hier fehlt einfach noch die Möglichkeit, konkret etwas vorzugeben, das Stable Diffusion dann auch verstehen und umsetzen kann.

Wenn du dein Ergebnis speichern möchtest, kannst du das Video als MP4-Datei herunterladen.

Dass künstliche Intelligenz auch Videos auf Basis konkreter Vorgaben erstellen kann, war eine Frage der Zeit – und Stable Video Diffusion ist sicherlich auch nicht die erste Anwendung dieser Art. Und doch ist das Ergebnis bereits beeindruckend. Die (nahe) Zukunft dürfte noch einige Verbesserungen bereithalten: Längere Videos mit besserer Steuerung, was in ihnen geschehen soll.