

# **Erstellung eines Trainingsdatensatzes für object detection mit echten und synthetischen Bildern**

Tim Tomczek  
Fachhochschule Südwestfalen

Schriftliche Ausarbeitung im Modul  
„Projekt KI“

22. April 2025

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>3</b>
1.1. Projektbeschreibung und Ziele . . . . .	3
1.2. Projekthintergrund . . . . .	3
1.3. Zeitplanung . . . . .	3
<b>2. Durchführung</b>	<b>3</b>
2.1. Recherche und Vergleich von Modellen . . . . .	3
2.2. Kurzvorstellung der Modelle . . . . .	4
2.3. Erstellung der Testumgebung . . . . .	4
2.4. Vergleich der Modelle . . . . .	4
<b>3. Anhang</b>	<b>7</b>
3.1. Bilder der Modellevaluation . . . . .	7

# 1. Einleitung

## 1.1. Projektbeschreibung und Ziele

## 1.2. Projekthintergrund

## 1.3. Zeitplanung

# 2. Durchführung

Den Start des Projekts bildet das Kickoff-Meeting mit dem Projektbetreuer und dem Projektdurchführenden am **29.03.2025**. In diesem Meeting wurde das Projekt vorgestellt und die Ziele sowie der Zeitplan besprochen. Anschließend wurde mit der Abgabe des Projektantrags das Projekt offiziell angemeldet.

## 2.1. Recherche und Vergleich von Modellen

Das Projekt beginnt mit der Recherche nach aktuellen KI-Modellen, die für die Generierung von Bildern geeignet sind. Dafür wurde zuerst nach entsprechend Benchmarks für die Vergleich von Modellen gesucht. Ein solcher Benchmark hätte es ermöglicht, die Modelle direkt miteinander zu vergleichen und die besten Modelle für das Projekt auszuwählen. Leider gibt es keinen allgemeingültigen Benchmark, der für alle Modelle geeignet ist. Daher wurde beschlossen, die vier obersten Modelle auf der HuggingFace-Website unter der Kategorie Text-To-Image, nach der Sortierung Trending, zu vergleichen. Dabei wurden verschiedene Versionen des gleichen Modells nicht berücksichtigt und jeweils das größte Modell ausgewählt. Fine-Tunings von Modellen wurden auch aus der Auswahl ausgeschlossen um einen möglichst repräsentativen Vergleich von verschiedenen Architekturen zu erhalten. Daraus entstand am 17.04.2025 die folgende Liste von Modellen, die für den Vergleich verwendet werden:

- HiDream-ai/HiDream-I1-Full
- black-forest-labs/FLUX.1-dev
- stabilityai/stable-diffusion-3.5-large
- stabilityai/stable-diffusion-xl-base-1.0

Diese Liste stellt die Modelle in der Reihenfolge dar, in der sie nach der Trending Sortierung angezeigt wurden.

## 2.2. Kurzvorstellung der Modelle

## 2.3. Erstellung der Testumgebung

## 2.4. Vergleich der Modelle

Für die Auswahl des Modells wurden verschiedenen Kriterien festgelegt, die für ein Modell welches den Anforderungen des Projekts gerecht werden soll, wichtig sind.

Das erste Kriterium ist die Anzahl der Fehler in den Bild. Für dieses Kriterium wurden die generierten Bilder durch den Autor auf visuelle Fehler untersucht. Als visueller Fehler wird dabei eine Abweichung von einem realen Bild der gleichen Szene definiert. Dies können Artefakte, falsche Anatomie von Personen oder Fehler in der Darstellung der Instrumente sein.

Das zweite Kriterium ist die Diversität der dargestellten Instrumente. Da in der nachfolgenden Bachelorarbeit ein Modell für Objekterkennung von Gitarren trainiert werden soll, ist es wichtig, dass das Modell verschiedene Gitarren darstellen kann, aber auch andere, ähnliche, Instrumente wie Geigen, Cellos, oder Banjos. Die Diversität wird dabei als Anzahl unterschiedlicher Instrumente und deren Variationen in den generierten Bildern definiert. Dafür wird die Anzahl der verschiedenen Instrumente in den generierten Bildern gezählt. Eine Variation eines Instruments wird mit einem halben Punkt bewertet und ein neues Instrument mit einem Punkt. Die Punkte werden dann addiert und ergeben die Diversität der Instrumente in den generierten Bildern.

Das dritte Kriterium ist die Diversität der Szene. Diese wird ähnlich wie die Diversität der Instrumente bewertet. Pro Modell werden Bilder mit zwei Eingaben generiert. Die Modelle verwenden dabei die empfohlenen Parameter, wobei pro Eingabe drei Bilder erstellt werden. Pro Bild variiert der *guidance\_scale* um den, vom Modellm, empfohlenen Wert plus zwei und minus zwei. Jede Szene wird dabei mit einem Punkt bewertet. Eine Variation einer bereits vorhandenen Szene wird mit einem halben Punkt bewertet.

Die Gesamtwertung pro Bild wird dann aus der Summe der Punkte der Diversitätskriterien abzüglich der Anzahl der Fehler in den Bildern gebildet. Für die Modellwertung werden dann die Einzelwertungen addiert und durch die Anzahl der Bilder geteilt. Das Modell mit der höchsten Gesamtwertung wird dann als das beste Modell ausgewählt.

In Tabelle 1 sind die Ergebnisse der Bewertung der Modelle dargestellt. Die untersuchten Bilder mit dne markierten Fehlern und Instrumenten sind in Anhang ?? zu finden.

Basierend auf den Ergebnissen aus Tabelle 1 ergeben sich die folgenden Gesamtwertungen für die Modelle:

- HiDream-I1-Full: -3
- FLUX-1-dev: 1,75

- sd35: 0,08
- sdxlrefiner: -6,1

Das Modell **FLUX-1-dev** hat die höchste Gesamtwertung und wird daher für die weitere Arbeit verwendet. Das Modell **sd35** hat eine ähnliche Gesamtwertung, jedoch ist die Diversität der Instrumente und Szenen geringer als bei dem FLUX-Modell. Daher wird das Modell **sd35** nicht weiter betrachtet. Die Modelle **HiDream-I1-Full** und **sdxlrefiner** haben eine deutlich schlechtere Gesamtwertung und werden daher nicht weiter betrachtet.

Dateiname	Anzahl Fehler	Diversität Instrumente	Diversität Szene	Wertung
HiDream-I1-Full_prompt-0_guidance-3.0_steps-50.png	7	3,5	1	-2,5
HiDream-I1-Full_prompt-0_guidance-5.0_steps-50.png	9	1,5	0,5	-7
HiDream-I1-Full_prompt-0_guidance-7.0_steps-50.png	8	4	0,5	-3,5
HiDream-I1-Full_prompt-1_guidance-3.0_steps-50.png	5	2	1	-2
HiDream-I1-Full_prompt-1_guidance-5.0_steps-50.png	5	1	0,5	-3,5
HiDream-I1-Full_prompt-1_guidance-7.0_steps-50.png	3	2,5	0,5	0
FLUX-1-dev_prompt-0_guidance-1.5_steps-50.png	2	2,5	1	1,5
FLUX-1-dev_prompt-0_guidance-3.5_steps-50.png	4	3	1	0
FLUX-1-dev_prompt-0_guidance-5.5_steps-50.png	3	4	1	2
FLUX-1-dev_prompt-1_guidance-1.5_steps-50.png	2	1	1	0
FLUX-1-dev_prompt-1_guidance-3.5_steps-50.png	1	3,5	1	3,5
FLUX-1-dev_prompt-1_guidance-5.5_steps-50.png	1	3,5	1	3,5
sd35_prompt-0_guidance-1.5_steps-28.png	2	1	1	0
sd35_prompt-0_guidance-3.5_steps-28.png	2	2,5	1	1,5
sd35_prompt-0_guidance-5.5_steps-28.png	4	3	1	0
sd35_prompt-1_guidance-1.5_steps-28.png	5	3	1	-1
sd35_prompt-1_guidance-3.5_steps-28.png	4	3,5	1	0,5
sd35_prompt-1_guidance-5.5_steps-28.png	5	2,5	1	-1,5
sdxlrefiner_prompt-0_guidance-5.5_steps-40.png	7	0	1	-6
sdxlrefiner_prompt-0_guidance-7.5_steps-40.png	7	0	1	-6
sdxlrefiner_prompt-0_guidance-9.5_steps-40.png	8	0	1	-7
sdxlrefiner_prompt-1_guidance-5.5_steps-40.png	7	0	1	-6
sdxlrefiner_prompt-1_guidance-7.5_steps-40.png	8	0	1	-7
sdxlrefiner_prompt-1_guidance-9.5_steps-40.png	7	1	1	-5

Tabelle 1: Bewertung der Modelle

### **3. Anhang**

#### **3.1. Bilder der Modellevaluation**

## **Erklärung**

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und dabei keine anderen als die angegebenen Hilfsmittel benutzt habe. Sämtliche Stellen der Arbeit, die im Wortlaut oder dem Sinn nach Werken anderer Autoren entnommen sind, habe ich als solche kenntlich gemacht. Die Arbeit wurde bisher weder gesamt noch in Teilen einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

22. April 2025

Tim Tomczek