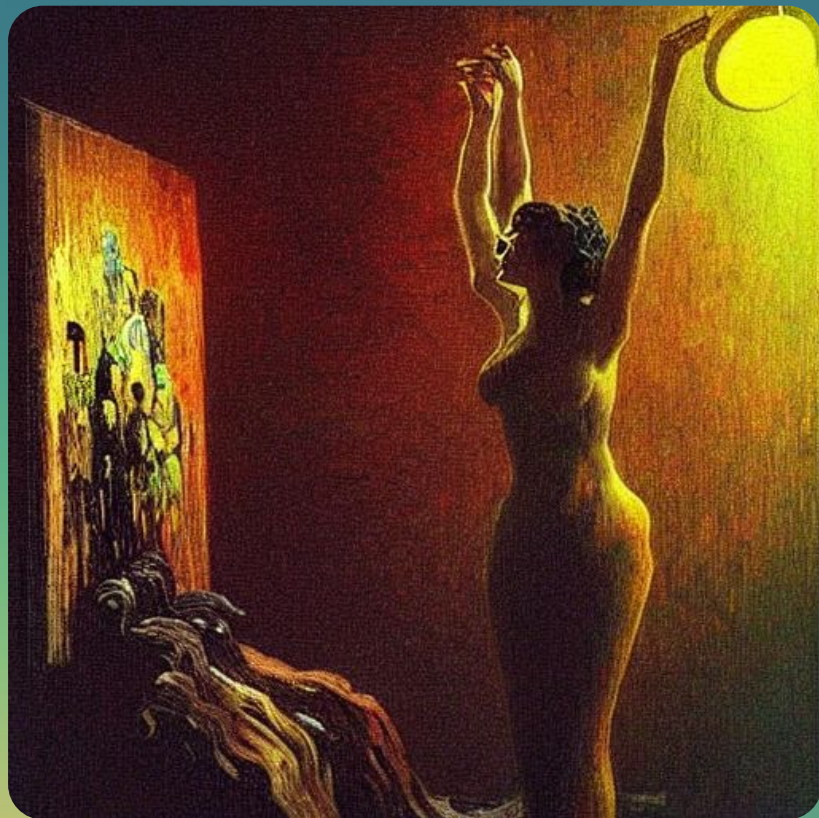# A.I. Music Videos

Automatische Erstellung von Musikvideos durch K.I.

ARIC Brownbag von Anton Wiehe
Contact at anton.wiehe@adalab.ai

Dieses Projekt entsteht in Kooperation mit

hauptsache.net

Hosted in Germany

GDPR Compliant

Made in Hamburg

Erweitertes Team

Angelie Kraft
Co-Founder, Full-time PHD

Yannic Jänike
Machine Learning Engineer

Anton Wiehe, Pia Čuk, Florian Woeste

Intelligent Adaptive Systems, M.Sc.

Besucht uns in unserem Office:
Neuer Pferdemarkt 1

Die Ideale Pipeline

1.
Riesiger Datensatz an Musikvideos

2.
Modell trainieren...

3.
???

4.
Profit

Das Problem:
Gibt es (noch) nicht.
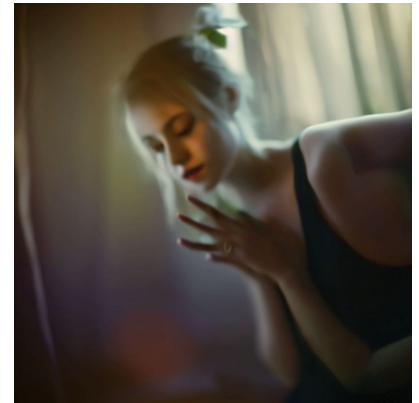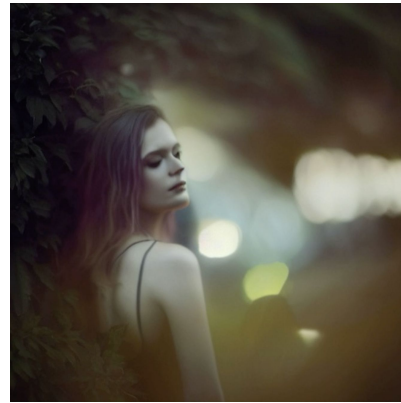Videos müssen getagged sein für Steuerung durch Künstler:innen

Die Lösung:
Es gibt Komponenten für die Schritte!
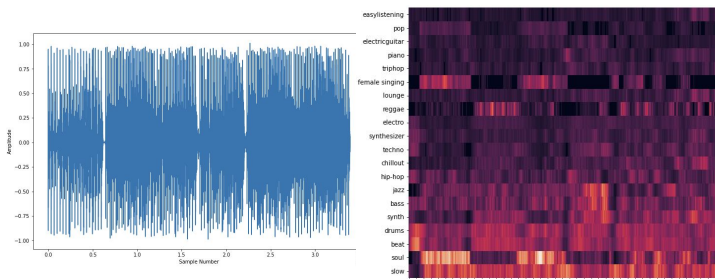Eine Kaskade an Netzwerken ermöglicht hohe Qualität und Steuerbarkeit.
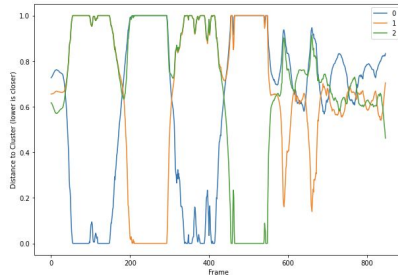
Beispielvideo

"Daughter of the sun"
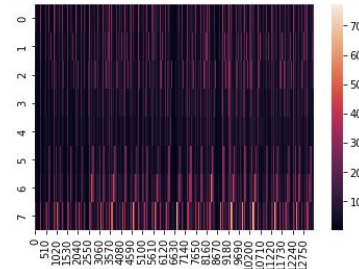von vbnd

Unsere Pipeline
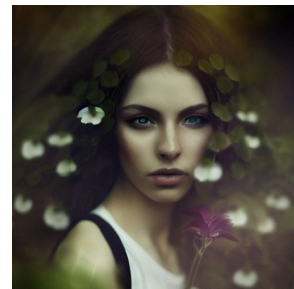

1
Song as raw waveform

2
Music tagging models create tags

3
The song is clustered by tags and scenes are described
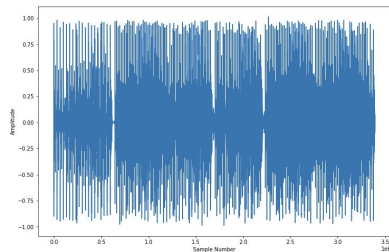
4
A spectrogram analyzes the tonality

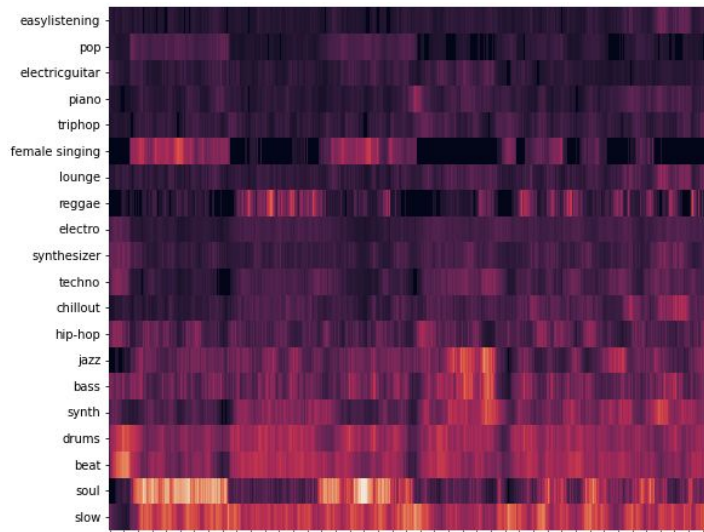**5**
Individual frames are generated

Unsere Pipeline:
"Music to Text"-Tagging + Clustering
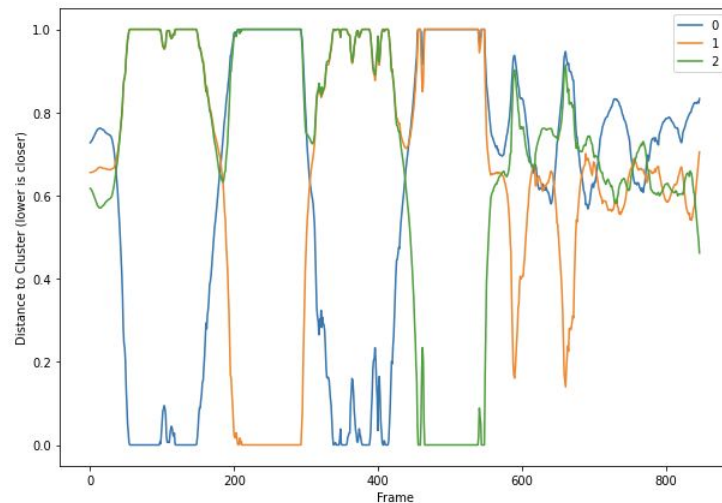
Tagging

Waveform

Taggram

Clustering

Cluster tags

1. soul, slow, drums, female singing

2. slow, beat, drums, synth

3. synth, jazz, drums, bass

Main tags slow, soul, beat, drums, synth

Prompting GPT to write complete scenes

We use: CLIP-guided sampling + Reranking

```
The goal is a vivid description of an image without mentioning music,
instruments, and music genres. First, adjectives describing the image
are given in the order of their importance. Then, a full description
of the corresponding imagery is given:
 sad, beautiful, soft, quiet, slow. Image description: An old widow
is sitting on a chair in a beautiful garden with her hands folded in
front of her. She is looking at you with a sad expression on her
face.
 electronic, loud, happy, abstract. Image description: Dynamic and
vibrant colors are forming strong geometric shapes that resemble a
rave.
 weird, happy, fast. Image description: A man is experiencing a
strange dream. He is struggling to feel his emotions as they rush too
quickly through his body. He is an a state of ecstacy.
 harmonious, mellow. Image description: An electric light begins to
dim at a distant point in the sky. You feel complete and at one with
your environment.
 soul, slow, drums, female singing, rnb, hip-hop. Image description:
```

Cluster tags

1. soul, slow, drums, female singing
2. slow, beat, drums, synth
3. synth, jazz, drums, bass

Main tags slow, soul, beat, drums, synth

GPT-generated cluster sentences:

1. "Flower music has become a common and somewhat trendy accompaniment for pop and rock music in the last decade. These songs typically blend soulful female vocals with drum machines, keyboards, and saxophones"

2. "Flower music and upbeat drums are layered into the background as you feel completely at ease in this house"

3. "Flower like sounds from electronic drums fill the air as a funky bass line slowly builds up in the background"

GPT-generated main artist "By Thomas Calder."

Text to Image - Stable Diffusion

Tags

'bass': 0.23
'quiet': 0.19
'synthesizer': 0.18
'hip-hop': 0.17
'techno': 0.14

Effects

Lyrics

Generated
Sentences

Stable Diffusion!

CLIP-encodings sind eigentlich eindimensionale Vektor der Größe (512)

Stable Diffusion benutzt aber cross-attention auf den eingebetteten Tokens der Größe (77, 768)

➥ Das erschwert den Prozess, aber erlaubt viele Möglichkeiten für das Mixen von Texten.

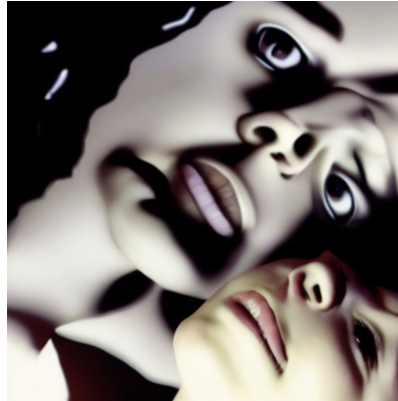**Unsere Pipeline:**
Image to Image - Interpolierung

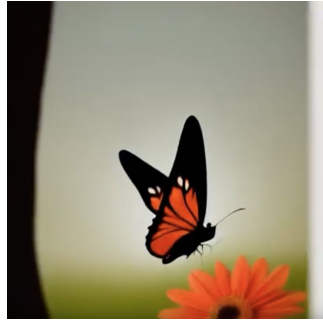Von 6 FPS zu über 60 FPS mit Interpolierungsnetzwerk

Beispielvideo
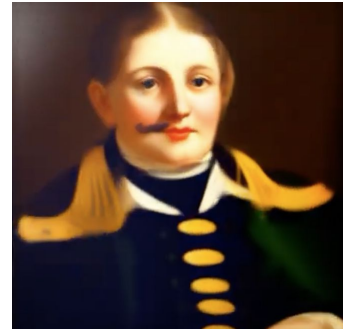
**"Shia Labeouf"
von Rob Cantor**

Mehr Beispiele

"Space is the place" von Men I trust:
https://youtu.be/PErzUfvKdcE

"Daughter of the Sun" von vbnd (Variation):
https://youtu.be/l-IUgAlqoHo

Danke für eure Aufmerksamkeit!

Ihr wollt mehr über K.I.-generierte Kunst wissen?
Oder ihr wollt selbst mit Kunst-generierender K.I. arbeiten?
Schreibt uns unter hello@adalab.ai oder anton.wiehe@adalab.ai