

Documentatie Tema 2 AM

Modelul generativ de imagini

Acest model a fost folosit cu scopul de a genera o imagine pe baza unui text prompt. Este setata un width si un height de 512 pixeli pentru fiecare poza generata. Modelul a fost rulat cu un batch_size de 3 astfel incat sa genereze un set de 3 poze. De asemenea, s-a folosit un prompt specific astfel incat imaginea generate sa fie suficient de realista si de simpla incat sa poata sa fie prelucrata ulterior Timpul de generare al pozelor a fost aproximativ ~1 hour.

Poza generata dupa rularea modelului: alt text

Sintetizare Voce:

Sintetizatorul de voce a fost folosit pentru a genera un fisier audio pe baza unui text prompt. A fost folosit un dictionar cu prompt-uri pentru fiecare fisier in parte care apare in videoul final. Fisierele au fost salvate in format .wav si au fost prelucrate ulterior pentru a fi adaugate in videoul final.

Spatiul de culoare:

Spati-ul de culoare folosit este YUV. Convertirea din BGR in YUV a fost facuta folosind functia `cv2.cvtColor()`. CV2 este o biblioteca de functii pentru procesarea imaginilor si a video-urilor. Aceasta biblioteca este folosita pentru a converti o imagine dintr-un spatiu de culoare in altul. In cazul nostru, imaginea a fost convertita din spatiul de culoare BGR in spatiul de culoare YUV intrucat `cv2` foloseste spatiul de culoare BGR ca default.

Video:

Pentru a crea un video, am folosit biblioteca `moviepy`. Aceasta biblioteca este folosita pentru a edita si a crea video-uri. Am folosit functia `ImageClip` pentru a adauga imagini in video si functia `TextClip` pentru a adauga text in video. De asemenea, am folosit functia `concatenate_videoclips` pentru a concatena clipurile video si functia `write_videofile` pentru a salva video-ul final. Pentru a concatena audio-urile am folosit functia `concatenate_audioclips` si dupa am asociat aceste audio-uri cu video-urile deja formate.

Mask Culorii:

Masca culorii a fost aleasa pentru a putea separa culoare obiectului (fuchsia) si background-ul (verde cu alb). Masca a fost creata folosind functia `cv2.inRange()` care returneaza o masca binara a imaginii. Masca a fost aplicata pe imaginea originala folosind functia `cv2.bitwise_and()`. Masca a fost folosita pentru a crea un contur in jurul obiectului si pentru a-l separa de background.

Rezultat final:

Rezultatul final este un video care contine o imagine generata de modelul generativ de imagini, un fisier audio generat de sintetizatorul de voce si un video final care le contine pe toate