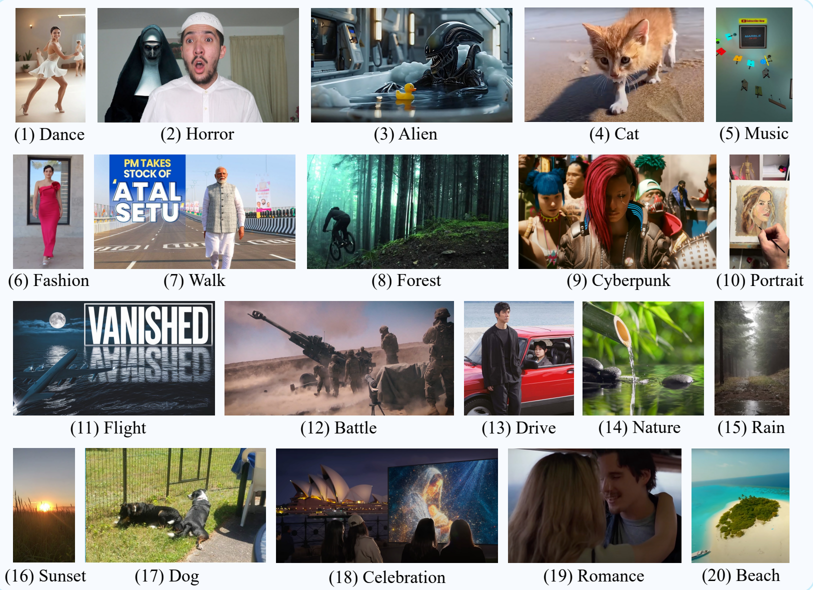
VideoUFO: A Million-Scale User-Focused Dataset for Text-to-Video Generation

Introduction



- We present VideoUFO, the first video dataset curated based on the focus of real text-to-video users. This dataset comprises over 1.09 million clips spanning 1, 291 user-focused topics.
- We compare VideoUFO with recent video datasets, highlighting their differences in both fundamental attributes and topics coverage, thereby emphasizing the necessity of our dataset. We also follow best practices in their curation processes to ensure the quality of our dataset.
- We evaluate current text-to-video models on user-focused topics and observe that a simple model trained on our VideoUFO outperforms competing models on worst-performing topics.

A Datapoint in VideoUFO

Video Clip

ID --7WJyWZf8A.0

Topic *Car*

Start Time *0:00:02.633*

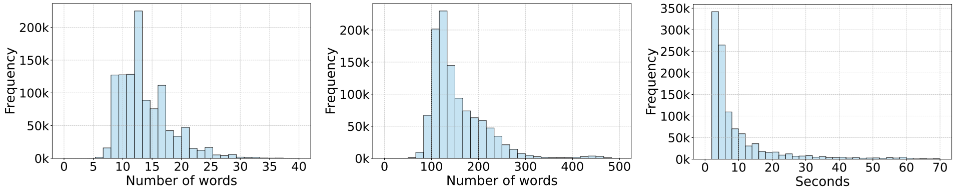
End Time *0:00:08.633*

Brief Caption: *A group of toy cars and a helicopter on a white backg-round.*

Detailed Caption: *The video begins with a white background featuring three animated vehicles: a police car, a fire truck, and an ambulance. The police car is black with blue lights on top, the fire truck is red with a yellow helmet on top, and the ambulance is white with green lights on top. Each vehicle has a smiling face with large eyes and a mouth, giving them a friendly and approachable appearance. As the video progresses, a red car with a bow on top enters the scene from the left side. This car has a smiling face and large eyes, similar to the other vehicles. It moves towards the center of the frame, where it stops and interacts with the other vehicles.*

Subject Consistency	Background Consistency	Motion Smoothness	Dynamic Degree	Aesthetic Quality	Imaging Quality
0.878	0.929	0.989	1.00	0.585	0.447

Statistics

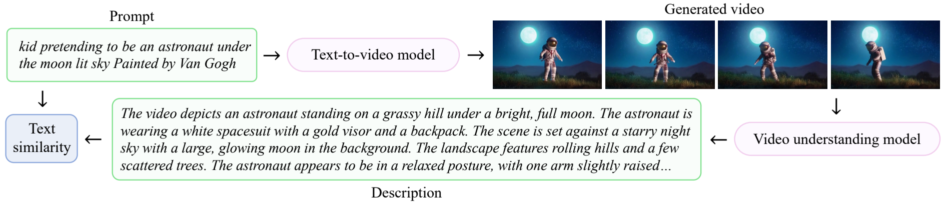


(a) Brief captions (b) Detailed captions (c) Clip duration

Comparison with other datasets

Dataset	#Vid.	$\overline{\text{Len.}}$	$\overline{\text{Words}}$	Resolution	Domain	#Topic	License
WebVid-10M [18]	10M	18.0s	14.2	<360p	Open	1,000	Retracted
HD-VILA-100M [11]	103M	13.4s	32.5	720p	Open	648	R-UDA
InternVid [16]	234M	11.7s	17.6	720p	Open	1,051	Apache 2.0
Panda-70M [16]	70M	8.5s	13.2	720p	Open	719	R-UDA
LVD-2M [14]	2M	20.2s	88.7	Diverse	Open	814	R-UDA
MiraData [15]	0.33M	72.1s	318.0	720p	Open	639	GPL 3.0
Koala-36M [13]	36M	17.2s	202.1	720p	Open	724	R-UDA
VidGen-1M [17]	1M	10.6s	89.3	720p	Open	835	R-UDA
OpenVid-1M [10]	1M	7.2s	127.3	Diverse	Open	671	R-UDA
VideoUFO	1M	12.6s	155.5	720p	**Users**	**1, 291**	**CC BY**

BenchUFO



Prompt

kid pretending to be an astronaut under the moon lit sky Painted by Van Gogh

→ Text-to-video model → Generated video

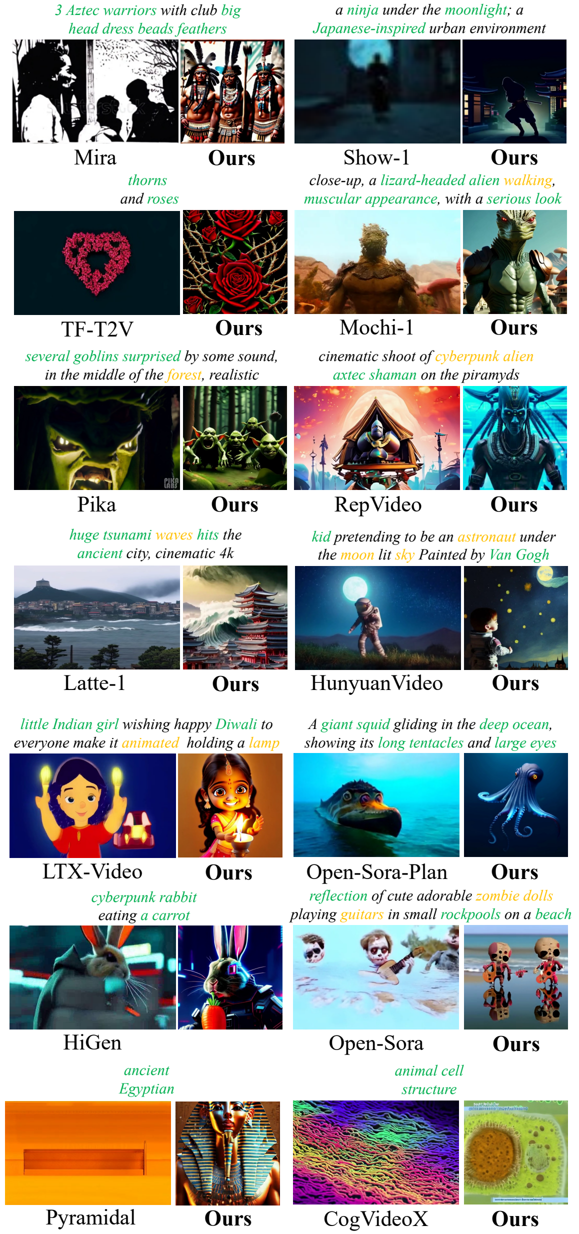
↓ Video understanding model ←

The video depicts an astronaut standing on a grassy hill under a bright, full moon. The astronaut is wearing a white spacesuit with a gold visor and a backpack. The scene is set against a starry night sky with a large, glowing moon in the background. The landscape features rolling hills and a few scattered trees. The astronaut appears to be in a relaxed posture, with one arm slightly raised…

Description

← Text similarity ↓

Comparison


3 Aztec warriors with club big head dress beads feathers
Mira
Ours
a ninja under the moonlight; a Japanese-inspired urban environment
Show-1
Ours
thorns and roses
TF-T2V
Ours
close-up, a lizard-headed alien walking, muscular appearance, with a serious look
Mochi-1
Ours
several goblins surprised by some sound, in the middle of the forest, realistic
Pika
Ours
cinematic shoot of cyberpunk alien axtec shaman on the piramyds
RepVideo
Ours
huge tsunami waves hits the ancient city, cinematic 4k
Latte-1
Ours
kid pretending to be an astronaut under the moon lit sky Painted by Van Gogh
HunyuanVideo
Ours
little Indian girl wishing happy Diwali to everyone make it animated holding a lamp
LTX-Video
Ours
A giant squid gliding in the deep ocean, showing its long tentacles and large eyes
Open-Sora-Plan
Ours
cyberpunk rabbit eating a carrot
HiGen
reflection of cute adorable zombie dolls playing guitars in small rockpools on a beach
Open-Sora
Ours
ancient Egyptian
Pyramidal
Ours
animal cell structure
CogVideoX
Ours