Locally Hosted LTX-Video Analysis

Executive Summary

This report focuses on the deployment and performance analysis of the Lightricks/LTX-Video model package, a promising open-source video generation tool. We investigate its out-of-the-box capabilities, demonstrating how to optimize its performance for practical applications without requiring extensive retraining or fine-tuning. A key challenge addressed is the extension of generated video length to 10 seconds, achieved by strategically increasing the number of frames while maintaining a consistent frame rate. This analysis reveals that the LTX-Video model is notably lightweight and memory-efficient, leading to faster video generation times.

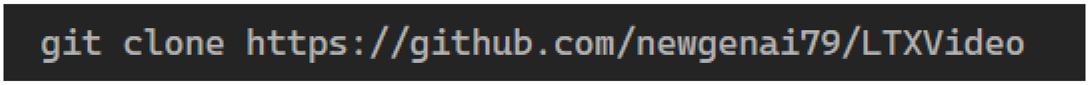
Running LTX-Video

Tested LTX-Video model running locally after cloning their repository and modifying some of the inference elements for convenience.
Without using int8 quantization, at inference model observed shared (GPU) memory usage of nearly 15GB initially and then it dropped down to 6GB.
While using the int8 quantization option, the GPU memory usage was around 4GB at inference.

The Process of Running LTX-Video

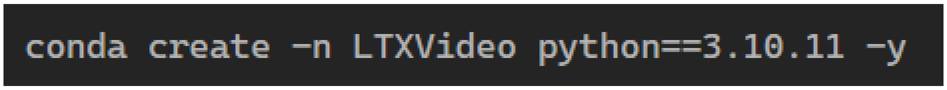
1. Clone the LTX-Video repository

    - I have used a slightly modified repository of the official repository to include some elements to make usage more convenient.
        - inference.py is modified to include a config file instead of getting the arguments from the terminal
        - A requirements.txt file is added to install dependencies



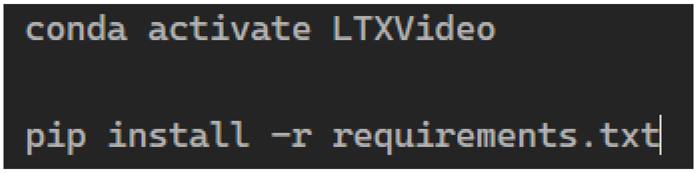
```
git clone https://github.com/newgenai79/LTXVideo
```

2. Create a virtual environment using conda



```
conda create -n LTXVideo python==3.10.11 -y
```

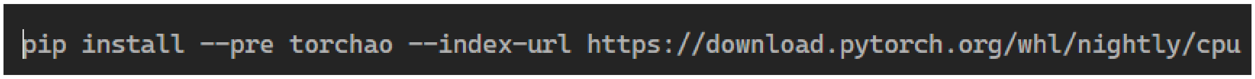
3. Activate the virtual environment and install the requirements file



```
conda activate LTXVideo

pip install -r requirements.txt
```

4. Install the torchon library

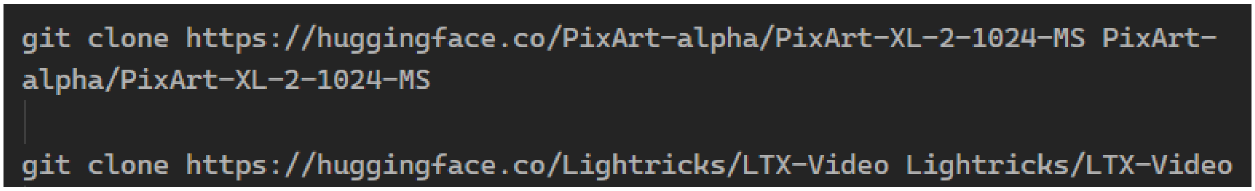
    - This library helps us to optimize and quantize the model



```
pip install --pre torchao --index-url https://download.pytorch.org/whl/nightly/cpu
```

5. Download the models

    - Clone the model repositories from huggingface



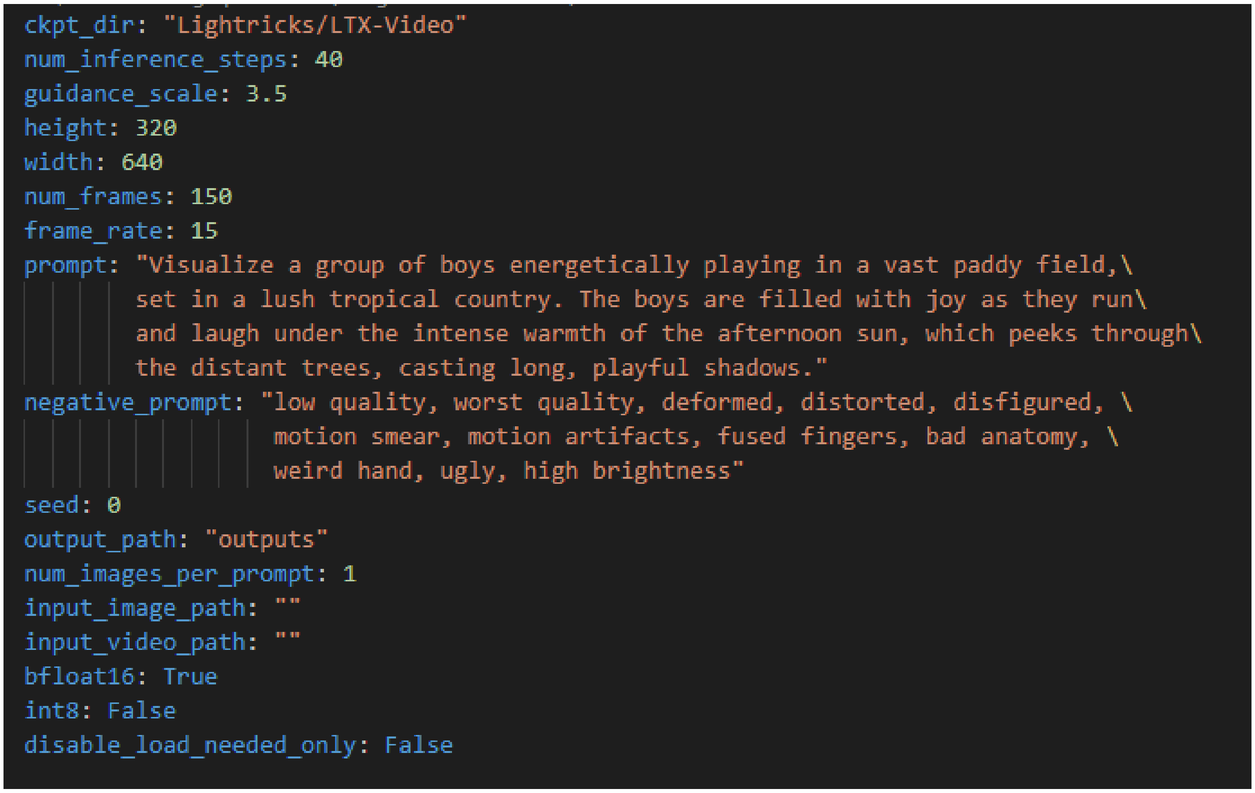
```
git clone https://huggingface.co/PixArt-alpha/PixArt-XL-2-1024-MS PixArt-alpha/PixArt-XL-2-1024-MS

git clone https://huggingface.co/Lightricks/LTX-Video Lightricks/LTX-Video
```

6. Modify the configuration file

    Following are some of the important configurations that can be done out of the box

    - num_inference_steps: The number of steps the model will take during the inference process. Higher values generally improve the quality of the generated video but at the cost of increased computation time.
    - guidance_scale: Adjusts the influence of the prompt on the generation process.
    - num_frames: The total number of frames in the generated video. Define this value according to the frame rate.
        - Eg: If you want a 10-sec clip and the frame rate is 15 fps, you need to define 150 as the value for this configuration
    - height: Can decrease to reduce GPU memory usage in exchange for lower resolution
    - width: Can also decrease to reduce GPU memory usage in exchange for lower resolution
    - prompt: The descriptive text that guides the video generation.

- negative_prompt: Ensures the generated content does not include undesirable elements.
- input_image_path: in addition to the prompt we can give an reference image from this configuration
- int8: If you make this ‘True’ model will run in ‘int8’ quantized mode saving VRAM
- disable_load_needed_only: This will also increase memory usage efficiency

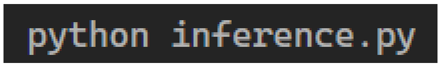


```yaml
ckpt_dir: "Lightricks/LTX-Video"
num_inference_steps: 40
guidance_scale: 3.5
height: 320
width: 640
num_frames: 150
frame_rate: 15
prompt: "Visualize a group of boys energetically playing in a vast paddy field,\
        set in a lush tropical country. The boys are filled with joy as they run\
        and laugh under the intense warmth of the afternoon sun, which peeks through\
        the distant trees, casting long, playful shadows."
negative_prompt: "low quality, worst quality, deformed, distorted, disfigured, \
                 motion smear, motion artifacts, fused fingers, bad anatomy, \
                 weird hand, ugly, high brightness"
seed: 0
output_path: "outputs"
num_images_per_prompt: 1
input_image_path: ""
input_video_path: ""
bfloat16: True
int8: False
disable_load_needed_only: False
```

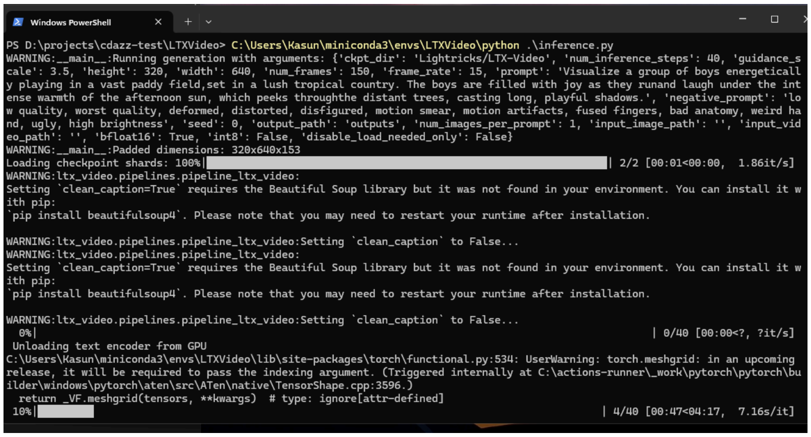
Example Configuration File

7. Running the inference

- Inference on the model by executing the inference.py python file.

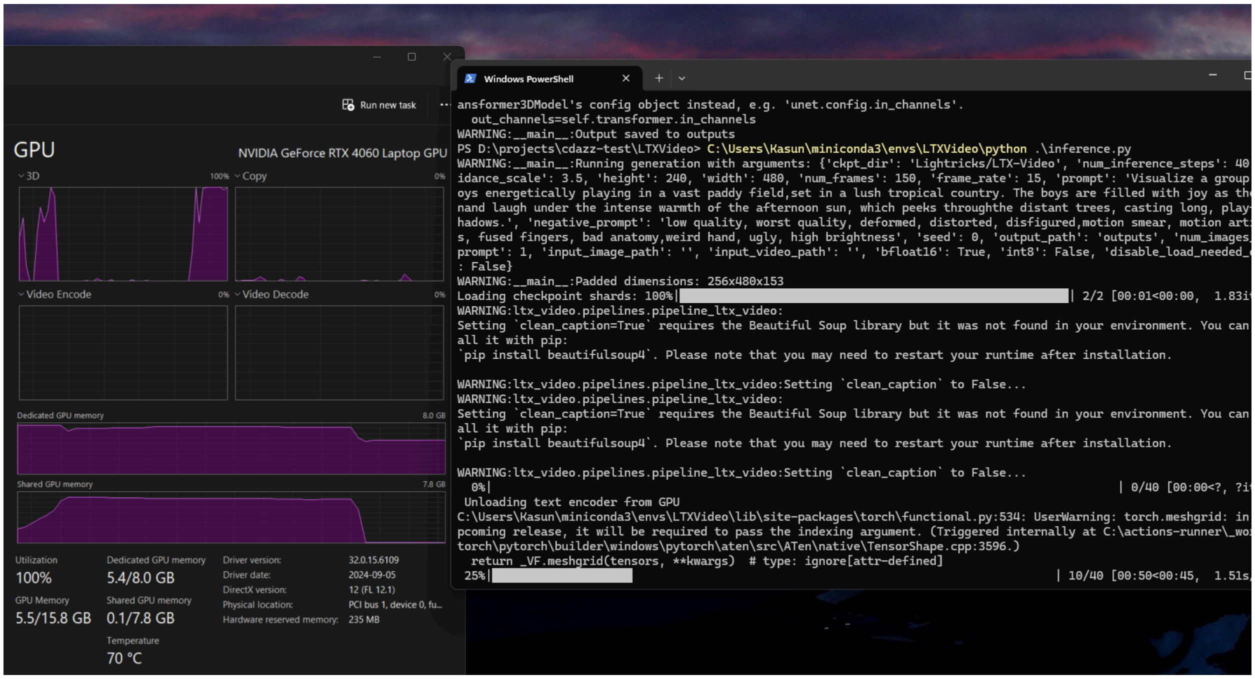


```
python inference.py
```



```
PS D:\projects\cdazz-test\LTXVideo> C:\Users\Kasun\miniconda3\envs\LTXVideo\python .\inference.py
WARNING:__main__:Running generation with arguments: {'ckpt_dir': 'Lightricks/LTX-Video', 'num_inference_steps': 40, 'guidance_s
cale': 3.5, 'height': 320, 'width': 640, 'num_frames': 150, 'frame_rate': 15, 'prompt': 'Visualize a group of boys energeticall
y playing in a vast paddy field,set in a lush tropical country. The boys are filled with joy as they runand laugh under the int
ense warmth of the afternoon sun, which peeks throughthe distant trees, casting long, playful shadows.', 'negative_prompt': 'lo
w quality, worst quality, deformed, distorted, disfigured, motion smear, motion artifacts, fused fingers, bad anatomy, weird ha
nd, ugly, high brightness', 'seed': 0, 'output_path': 'outputs', 'num_images_per_prompt': 1, 'input_image_path': '', 'input_vid
eo_path': '', 'bfloat16': True, 'int8': False, 'disable_load_needed_only': False}
WARNING:__main__:Padded dimensions: 320x640x153
Loading checkpoint shards: 100%|████████████████████████████████████████████████████████████████| 2/2 [00:01<00:00,  1.86it/s]
WARNING:ltx_video.pipelines.pipeline_ltx_video:
Setting `clean_caption=True` requires the Beautiful Soup library but it was not found in your environment. You can install it w
ith pip:
`pip install beautifulsoup4`. Please note that you may need to restart your runtime after installation.

WARNING:ltx_video.pipelines.pipeline_ltx_video:Setting `clean_caption` to False...
WARNING:ltx_video.pipelines.pipeline_ltx_video:
Setting `clean_caption=True` requires the Beautiful Soup library but it was not found in your environment. You can install it w
ith pip:
`pip install beautifulsoup4`. Please note that you may need to restart your runtime after installation.

WARNING:ltx_video.pipelines.pipeline_ltx_video:Setting `clean_caption` to False...
  0%|                                                                                          | 0/40 [00:00<?, ?it/s]
 Unloading text encoder from GPU
C:\Users\Kasun\miniconda3\envs\LTXVideo\lib\site-packages\torch\functional.py:534: UserWarning: torch.meshgrid: in an upcoming
release, it will be required to pass the indexing argument. (Triggered internally at C:\actions-runner\_work\pytorch\pytorch\bu
ilder\windows\pytorch\aten\src\ATen\native\TensorShape.cpp:3596.)
  return _VF.meshgrid(tensors, **kwargs)  # type: ignore[attr-defined]
 10%|█████                                                                          | 4/40 [00:47<04:17,  7.16s/it]
```

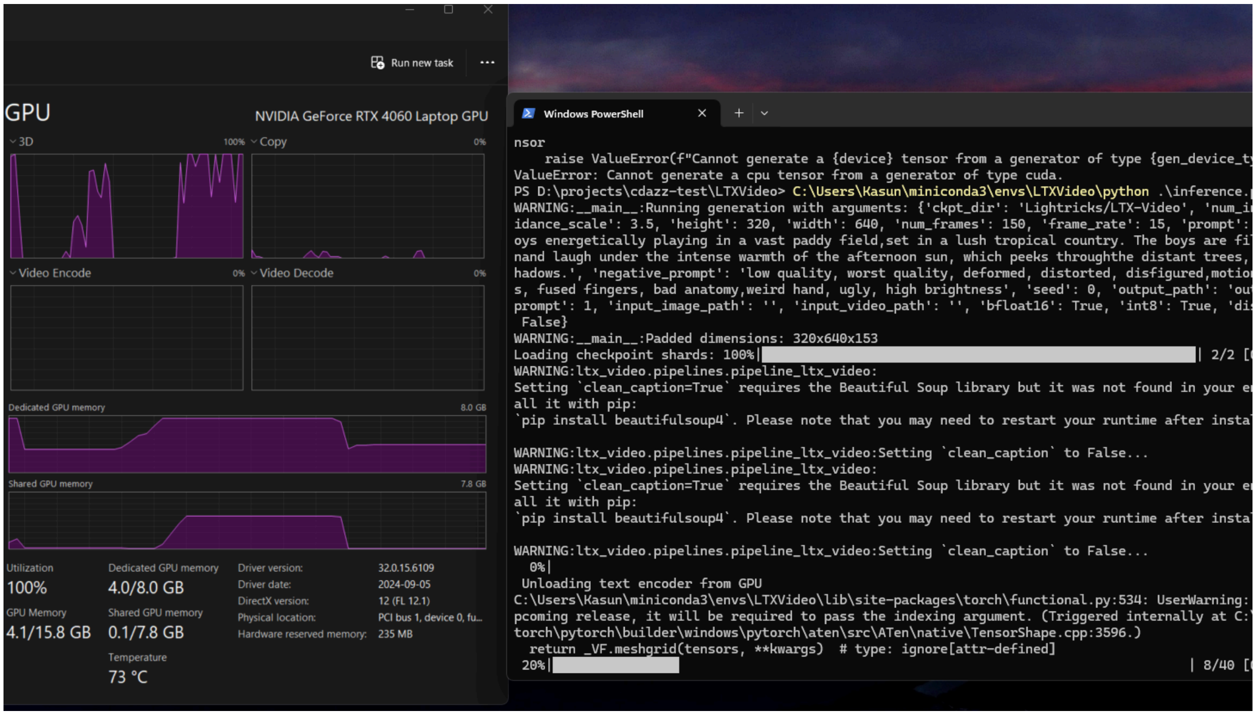
- As you can see from the below figure, at inference GPU memory usage was not over 6GB



Utilization 100%
Dedicated GPU memory 5.4/8.0 GB
GPU Memory 5.5/15.8 GB
Shared GPU memory 0.1/7.8 GB
Temperature 70 °C
Driver version: 32.0.15.6109
Driver date: 2024-09-05
DirectX version: 12 (FL 12.1)
Physical location: PCI bus 1, device 0, fu...
Hardware reserved memory: 235 MB

```
ansformer3DModel's config object instead, e.g. 'unet.config.in_channels'.
  out_channels=self.transformer.in_channels
WARNING:__main__:Output saved to outputs
PS D:\projects\cdazz-test\LTXVideo> C:\Users\Kasun\miniconda3\envs\LTXVideo\python .\inference.py
WARNING:__main__:Running generation with arguments: {'ckpt_dir': 'Lightricks/LTX-Video', 'num_inference_steps': 40
idance_scale': 3.5, 'height': 240, 'width': 480, 'num_frames': 150, 'frame_rate': 15, 'prompt': 'Visualize a group
oys energetically playing in a vast paddy field,set in a lush tropical country. The boys are filled with joy as the
nand laugh under the intense warmth of the afternoon sun, which peeks throughthe distant trees, casting long, play-
hadows.', 'negative_prompt': 'low quality, worst quality, deformed, distorted, disfigured,motion smear, motion arti
s, fused fingers, bad anatomy,weird hand, ugly, high brightness', 'seed': 0, 'output_path': 'outputs', 'num_images_
prompt': 1, 'input_image_path': '', 'input_video_path': '', 'bfloat16': True, 'int8': False, 'disable_load_needed_
: False}
WARNING:__main__:Padded dimensions: 256x480x153
Loading checkpoint shards: 100%|██████████████████████████████| 2/2 [00:01<00:00,  1.83it
WARNING:ltx_video.pipelines.pipeline_ltx_video:
Setting `clean_caption=True` requires the Beautiful Soup library but it was not found in your environment. You can
all it with pip:
`pip install beautifulsoup4`. Please note that you may need to restart your runtime after installation.

WARNING:ltx_video.pipelines.pipeline_ltx_video:Setting `clean_caption` to False...
WARNING:ltx_video.pipelines.pipeline_ltx_video:
Setting `clean_caption=True` requires the Beautiful Soup library but it was not found in your environment. You can
all it with pip:
`pip install beautifulsoup4`. Please note that you may need to restart your runtime after installation.

WARNING:ltx_video.pipelines.pipeline_ltx_video:Setting `clean_caption` to False...
  0%|                                                                            | 0/40 [00:00<?, ?it
 Unloading text encoder from GPU
C:\Users\Kasun\miniconda3\envs\LTXVideo\lib\site-packages\torch\functional.py:534: UserWarning: torch.meshgrid: in
pcoming release, it will be required to pass the indexing argument. (Triggered internally at C:\actions-runner\_wor
torch\pytorch\builder\windows\pytorch\aten\src\ATen\native\TensorShape.cpp:3596.)
  return _VF.meshgrid(tensors, **kwargs)  # type: ignore[attr-defined]
 25%|███████████                                                         | 10/40 [00:50<00:45,  1.51s/
```

- In the int8 quantization mode, the memory usage was around 4GB

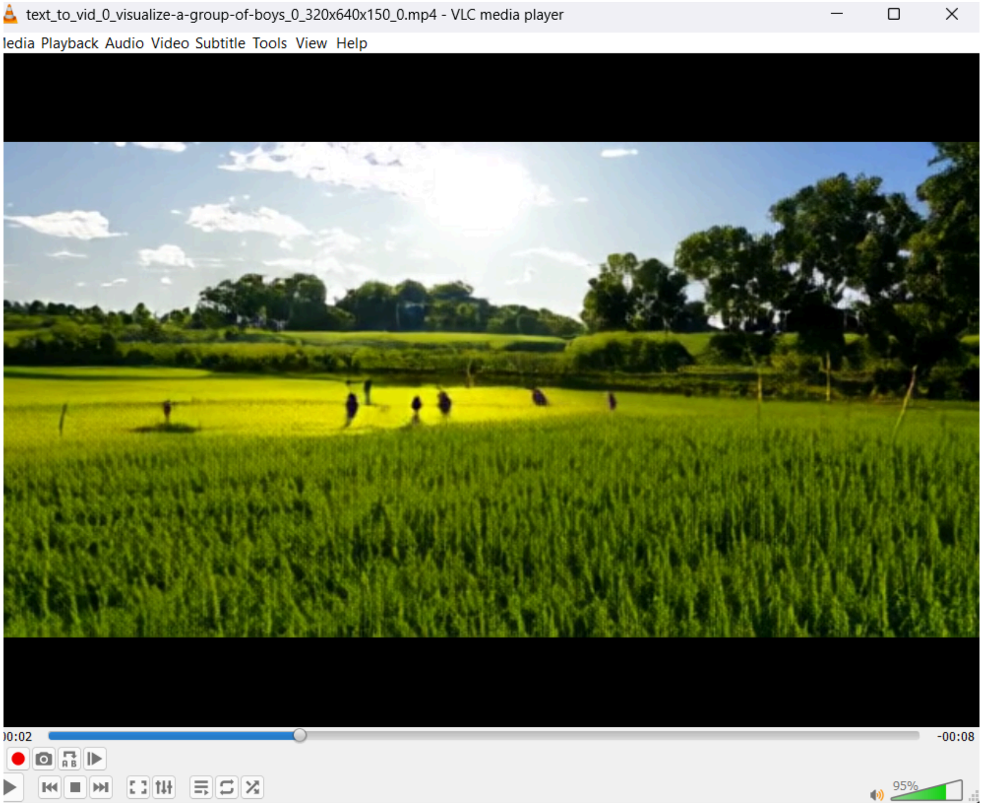


Utilization 100%
Dedicated GPU memory 4.0/8.0 GB
GPU Memory 4.1/15.8 GB
Shared GPU memory 0.1/7.8 GB
Temperature 73 °C
Driver version: 32.0.15.6109
Driver date: 2024-09-05
DirectX version: 12 (FL 12.1)
Physical location: PCI bus 1, device 0, fu...
Hardware reserved memory: 235 MB

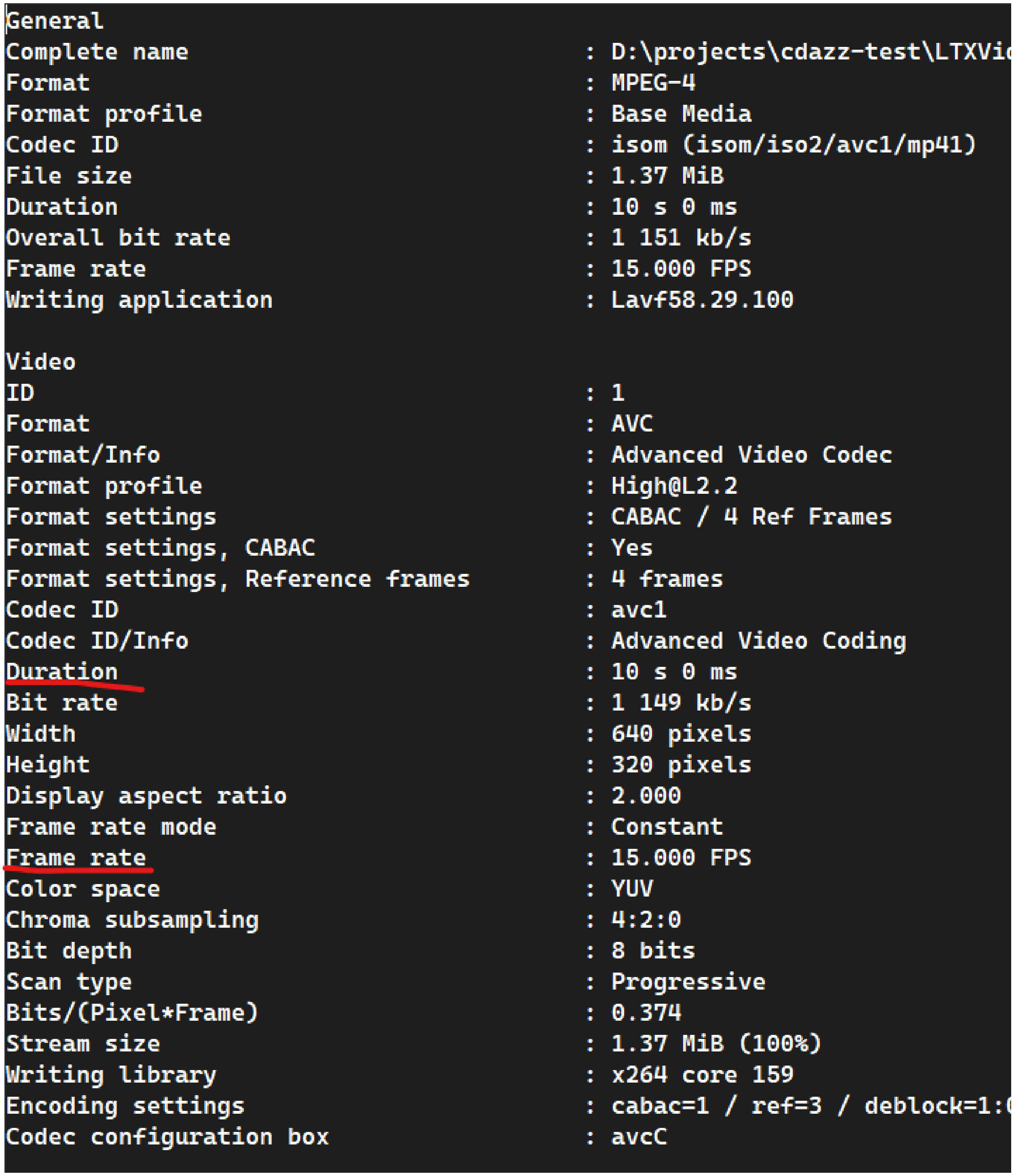
```
nsor
    raise ValueError(f"Cannot generate a {device} tensor from a generator of type {gen_device_ty
ValueError: Cannot generate a cpu tensor from a generator of type cuda.
PS D:\projects\cdazz-test\LTXVideo> C:\Users\Kasun\miniconda3\envs\LTXVideo\python .\inference.p
WARNING:__main__:Running generation with arguments: {'ckpt_dir': 'Lightricks/LTX-Video', 'num_in
idance_scale': 3.5, 'height': 320, 'width': 640, 'num_frames': 150, 'frame_rate': 15, 'prompt':
oys energetically playing in a vast paddy field,set in a lush tropical country. The boys are fil
nand laugh under the intense warmth of the afternoon sun, which peeks throughthe distant trees,
hadows.', 'negative_prompt': 'low quality, worst quality, deformed, distorted, disfigured,motion
s, fused fingers, bad anatomy,weird hand, ugly, high brightness', 'seed': 0, 'output_path': 'out
prompt': 1, 'input_image_path': '', 'input_video_path': '', 'bfloat16': True, 'int8': True, 'dis
 False}
WARNING:__main__:Padded dimensions: 320x640x153
Loading checkpoint shards: 100%|██████████████████████████████| 2/2 [0
WARNING:ltx_video.pipelines.pipeline_ltx_video:
Setting `clean_caption=True` requires the Beautiful Soup library but it was not found in your en
all it with pip:
`pip install beautifulsoup4`. Please note that you may need to restart your runtime after insta

WARNING:ltx_video.pipelines.pipeline_ltx_video:Setting `clean_caption` to False...
WARNING:ltx_video.pipelines.pipeline_ltx_video:
Setting `clean_caption=True` requires the Beautiful Soup library but it was not found in your en
all it with pip:
`pip install beautifulsoup4`. Please note that you may need to restart your runtime after insta

WARNING:ltx_video.pipelines.pipeline_ltx_video:Setting `clean_caption` to False...
  0%|
 Unloading text encoder from GPU
C:\Users\Kasun\miniconda3\envs\LTXVideo\lib\site-packages\torch\functional.py:534: UserWarning:
pcoming release, it will be required to pass the indexing argument. (Triggered internally at C:\
torch\pytorch\builder\windows\pytorch\aten\src\ATen\native\TensorShape.cpp:3596.)
  return _VF.meshgrid(tensors, **kwargs)  # type: ignore[attr-defined]
 20%|████████                                                          | 8/40 [0
```

8. Resulted Videos

- LTX-Video resultant videos are somewhat less than what we expected them to be as far as I experimented. If we manage to prompt engineer more in detail, it is possible that we can increase the response accuracy.


text_to_vid_0_visualize-a-group-of-boys_0_320x640x150_0.mp4 - VLC media player
ledia Playback Audio Video Subtitle Tools View Help
)0:02
-00:08
95%

- We can further see from the below video information the length of the video is 10 seconds and the frame rate is 15 fps



```
General
Complete name                            : D:\projects\cdazz-test\LTXVi
Format                                   : MPEG-4
Format profile                           : Base Media
Codec ID                                 : isom (isom/iso2/avc1/mp41)
File size                                : 1.37 MiB
Duration                                 : 10 s 0 ms
Overall bit rate                         : 1 151 kb/s
Frame rate                               : 15.000 FPS
Writing application                      : Lavf58.29.100

Video
ID                                       : 1
Format                                   : AVC
Format/Info                              : Advanced Video Codec
Format profile                           : High@L2.2
Format settings                          : CABAC / 4 Ref Frames
Format settings, CABAC                   : Yes
Format settings, Reference frames        : 4 frames
Codec ID                                 : avc1
Codec ID/Info                            : Advanced Video Coding
Duration                                 : 10 s 0 ms
Bit rate                                 : 1 149 kb/s
Width                                    : 640 pixels
Height                                   : 320 pixels
Display aspect ratio                     : 2.000
Frame rate mode                          : Constant
Frame rate                               : 15.000 FPS
Color space                              : YUV
Chroma subsampling                       : 4:2:0
Bit depth                                : 8 bits
Scan type                                : Progressive
Bits/(Pixel*Frame)                       : 0.374
Stream size                              : 1.37 MiB (100%)
Writing library                          : x264 core 159
Encoding settings                        : cabac=1 / ref=3 / deblock=1:
Codec configuration box                  : avcC
```

GitHub Link: https://github.com/RepZ97/text-to-video-gen-ltxvideo

Deployment

- Both of these model package repositories are open source and written in Python, making them suitable for deployment as Gunicorn servers once packaged with Docker.

- Then upon receiving an API call to make the inference, the docker application can do the inference and save the resultant video to an S3 bucket or send it to another model to upscale it as mentioned in the architecture diagram.

- As these model packages are intended to be deployed as part of the overall pipeline, no application user interface for inference was developed as a part of the report.

Annex

1. **GitHub Link:** https://github.com/RepZ97/text-to-video-gen-ltxvideo
2. **URL** for the Kaggle notebook: https://www.kaggle.com/code/kasundissanayake6962/cogvideox-kasun2