Blip Machine Setup

1. Create a machine on the Google Colud Platform with 20 N1 vCPUs and 120GB memory and a T4 GPU, using 1000GB Deep learning image CUDA11.8 disk
2. Run the following commands
   1. pip install accelerate
   2. pip install transformers
   3. Restart the machine
   4. pip install bitsandbytes

MusicGen Machine Setup

1. Using colab with V100 high RAM GPU
2. Run the following commands:
   1. python3 -m pip install setuptools wheel
   2. python3 -m pip install -U audiocraft
   3. sudo apt-get install ffmpeg
   4. python3 -m pip install torchvision
   5. python3 -m pip install --upgrade torch
   6. python3 -m pip install --upgrade torchaudio
3. git clone <https://github.com/facebookresearch/audiocraft.git>
4. cd audiocraft
5. python -m demos.musicgen\_app --share

Note that colab had an update recently. If the installation command doesn’t work, please use the following commands

* 1. python3 -m pip install setuptools wheel
  2. sudo apt-get install ffmpeg
  3. python3 -m pip install --upgrade torch
  4. python3 -m pip install torchvision
  5. python3 -m pip install --upgrade torchaudio
  6. python -m pip install -e . (need to clone the repository in your machine first)

Video segmentation machine

Simply follow the instructions on the website (https://www.scenedetect.com/), and it should work (we used Ubuntu22 environment for this final project

Short Clips:

1. Using the blip machine we setup
   1. Change the file name for the input video in the test2\_blip.py
   2. Run the test2\_blip.py to generate the video description
2. Using ChatGPT and type
   1. “Given <description> , generate an idea for music of it (in <num> words)
3. Using the web link generated by the program
   1. Copy the generated output from the ChatGPT and use the

Drama:

1. Using the scene segmentation environment
   1. Run “scenedetect --input orange.mp4 detect-adaptive list-scenes save-images -o out\_orange3”
   2. Take orange-Scenes.csv as the input to determine the input frames for the Blip2 model and the duration of the music generated by the MusicGen model
2. Using the Blip machine we set up
   1. Run the test\_complete\_orange.py to generate the test
3. Using the MusicGen environment we set up
   1. Take the text description as the input and set the duration of the generated music according to the configuration we get from the scene segmentation tool
4. Using colab environment
   1. Run try\_merge\_video\_and\_audioes.ipynb to merge the clips (using MoviePy)

Video Matching:

1. Upload the generated music or downloaded music
2. Run merge\_video\_and\_audio\_clip.ipynb on the colab with V100 GPU
   1. For content description matching, please run the merge\_video\_and\_audio\_clip\_content\_version.ipynb. However, those files are very similar