Image Generation in the Real World

I wanted to make digital/animated music videos for my music without spending money.

What I needed in an image generation tool:

- FREE!
- Video to video
- User friendly
- Smooth/not flickery

Warpfusion
https://github.com/Sxela/WarpFusion
Stable_warpfusion.ipynb
https://www.patreon.com/sxela

Free ❌✅ (mostly free)
Video to video ✅
User friendly ✅
Visually Smooth vid2vid ✅

- Based on disco diffusion
  - http://discodiffusion.com/
- Optional to support developer on Patreon
- Provides updates to warpfusion to include new features such as controlnets, loras, stable diffusion XL
- Tech support
- Discord community

Problems I ran into

- Google Colab became expensive
- Current hardware was too slow or could not handle image generation I wanted to do
- Initial set up can be difficult
  - I had trouble installing CUDAs/ setting up paths for cuda and python

Success!

clips of music videos
Full music video for my music

How to use warpfusion

https://github.com/Sxela/WarpFusion
Stable_warpfusion.ipynb
https://www.patreon.com/sxela

Basic Settings:

batch_name: " stable_warpfusion_0.13.0

Specify desired output size here.

Don't forget to rerun all steps after changing the width height (including forcing op

width_height: [1280,720]

Video Input Settings:

video_init_path: " video.mp4

extract_nth_frame: 1

define SD + K functions, load model

specify path to your Stable Diffusion checkpoint (the "original" flavor)

model_version: control_multi

vae_ckpt: " Insert text here

load_to: gpu

quantize: ☑

no_half_vae: ☐

model_path: " /content/drive/MyDrive/models/protogenV22Anime_22.safetensors

---

ControlNet download settings ControlNet downloads are managed by controlnet_multi settings in Main settings tab.

force_download: ☐

controlnet_models_dir: " /ControlNet

Seed and grad Settings:

set_seed: -1

Clamp grad is used with any of the init_scales or sat_scale above 0

Clamp grad limits the amount various criterions, controlled by *_scale pa

For example, high scale values may cause artifacts, and clamp_grad rem

clamp_grad: ☑

clamp_max: 2

Prompts

`animation_mode: None` will only use the first set. `animation_mode: 2D / Video` will run through them per the set fram
last one.

```
text_prompts = {0: ['a beautiful highly detailed cyberpunk mechanical \
augmented most beautiful (woman) ever, cyberpunk 2077, neon, dystopian, \
hightech, trending on artstation']}

negative_prompts = {
    0: ["text, naked, nude, logo, cropped, two heads, four arms, lazy eye, blurry, unfocused"]
}
```

Main settings.

Duplicated in the GUI and can be loaded there.

```
# DD-style losses, renders 2 times slower (!) and more memory intensive :D

latent_scale_schedule = [0,0] #controls coherency with previous frame in latent spa
init_scale_schedule = [0,0] #controls coherency with previous frame in pixel space.
sat_scale = 0

init_grad = False #True - compare result to real frame, False - to stylized frame
grad_denoised = True #fastest, on by default, calc grad towards denoised x instead
```

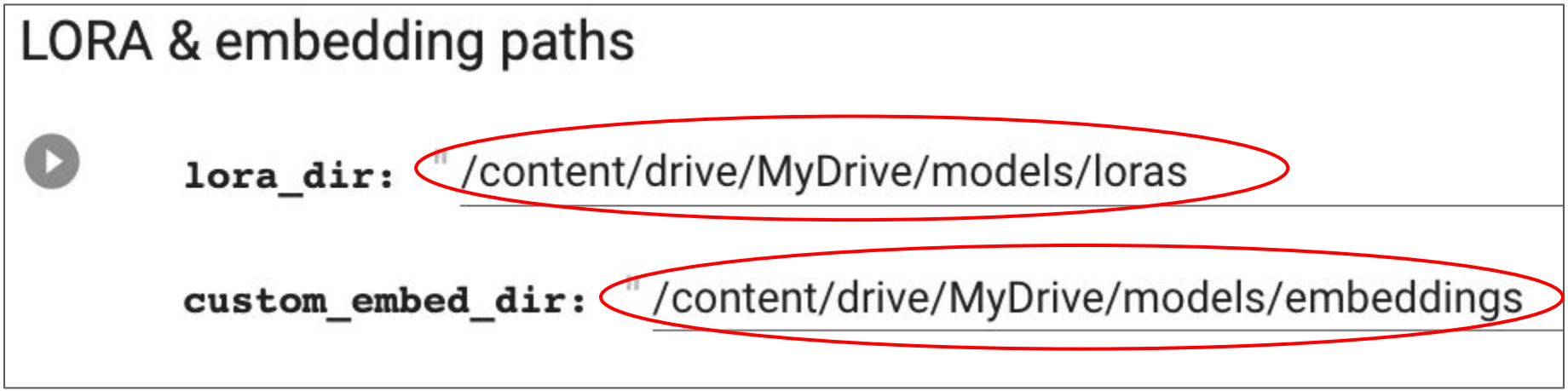
```
steps_schedule = {
    0: 25
} #schedules total steps. useful with low strength, when you end up with only 10 st
style_strength_schedule = [0.7]#[0.5]+[0.2]*149+[0.3]*3+[0.2] #use this instead of
flow_blend_schedule = [0.8] #for example [0.1]*3+[0.999]*18+[0.3] will fade-in for
cfg_scale_schedule = [15] #text2image strength, 7.5 is a good default
blend_json_schedules = True #True - interpolate values between keyframes. False - u

dynamic_thresh = 30
```

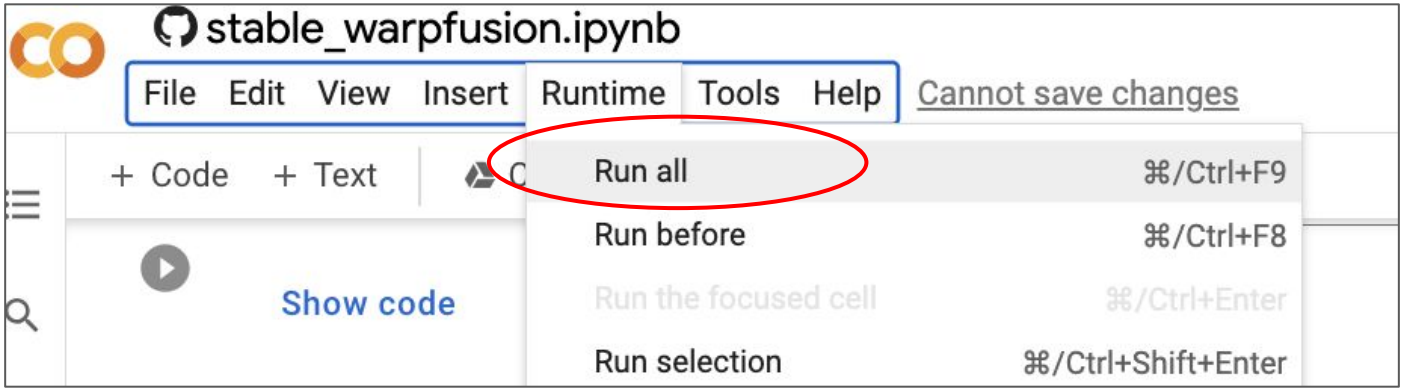
Style strength:

The higher the number, the more it follows the prompt

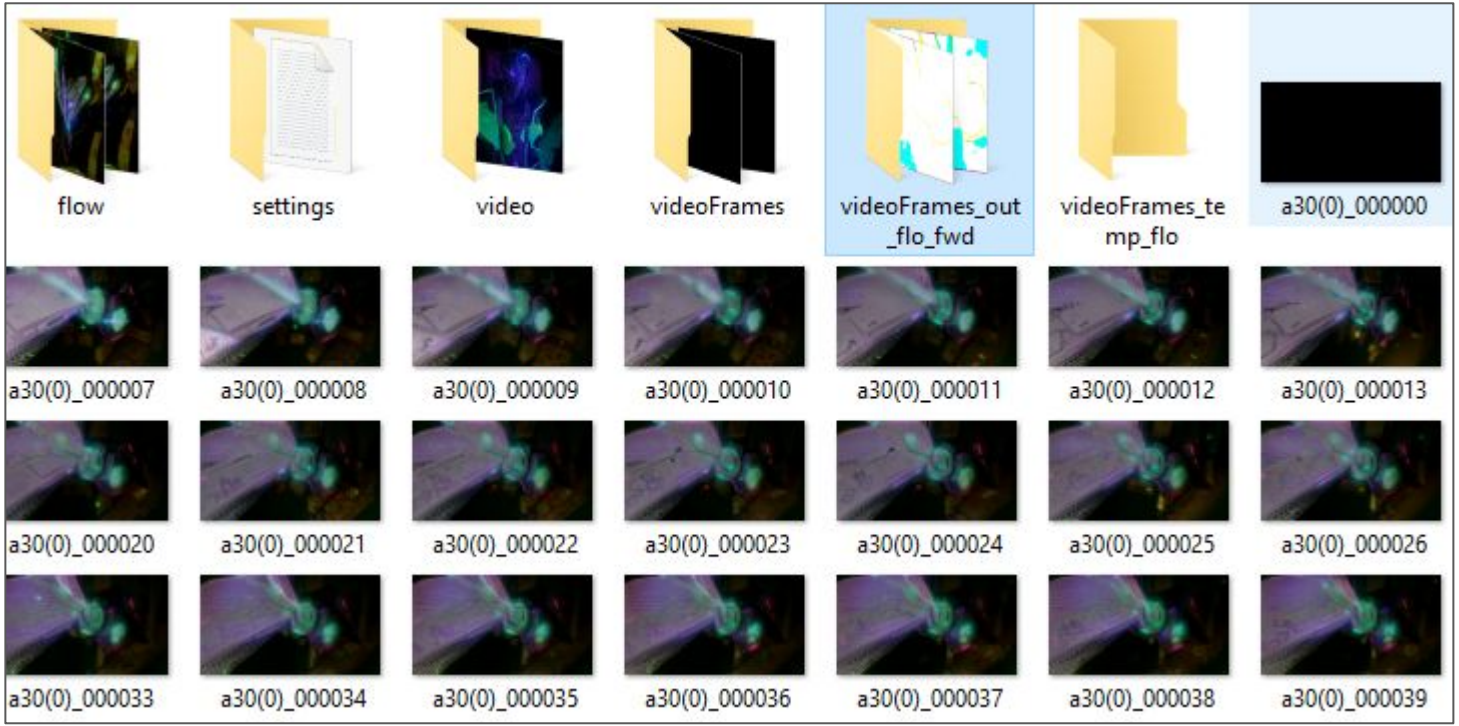
The lower the number, the more it resembles the original video

```
controlnet_multimodel = {
  "control_sd15_depth": {
    "weight": 0,
    "start": 0,
    "end": 1
  },
  "control_sd15_canny": {
    "weight": 0,
    "start": 0,
    "end": 1
  },
  "control_sd15_softedge": {
    "weight": 1,
    "start": 0,
    "end": 1
  },
  "control_sd15_mlsd": {
    "weight": 0,
    "start": 0,
    "end": 1
  },
  "control_sd15_normalbae": {
    "weight": 1,
    "start": 0,
    "end": 1
  },
  "control_sd15_openpose": {
    "weight": 1,
    "start": 0,
    "end": 1
  },
  "control_sd15_scribble": {
    "weight": 0
```

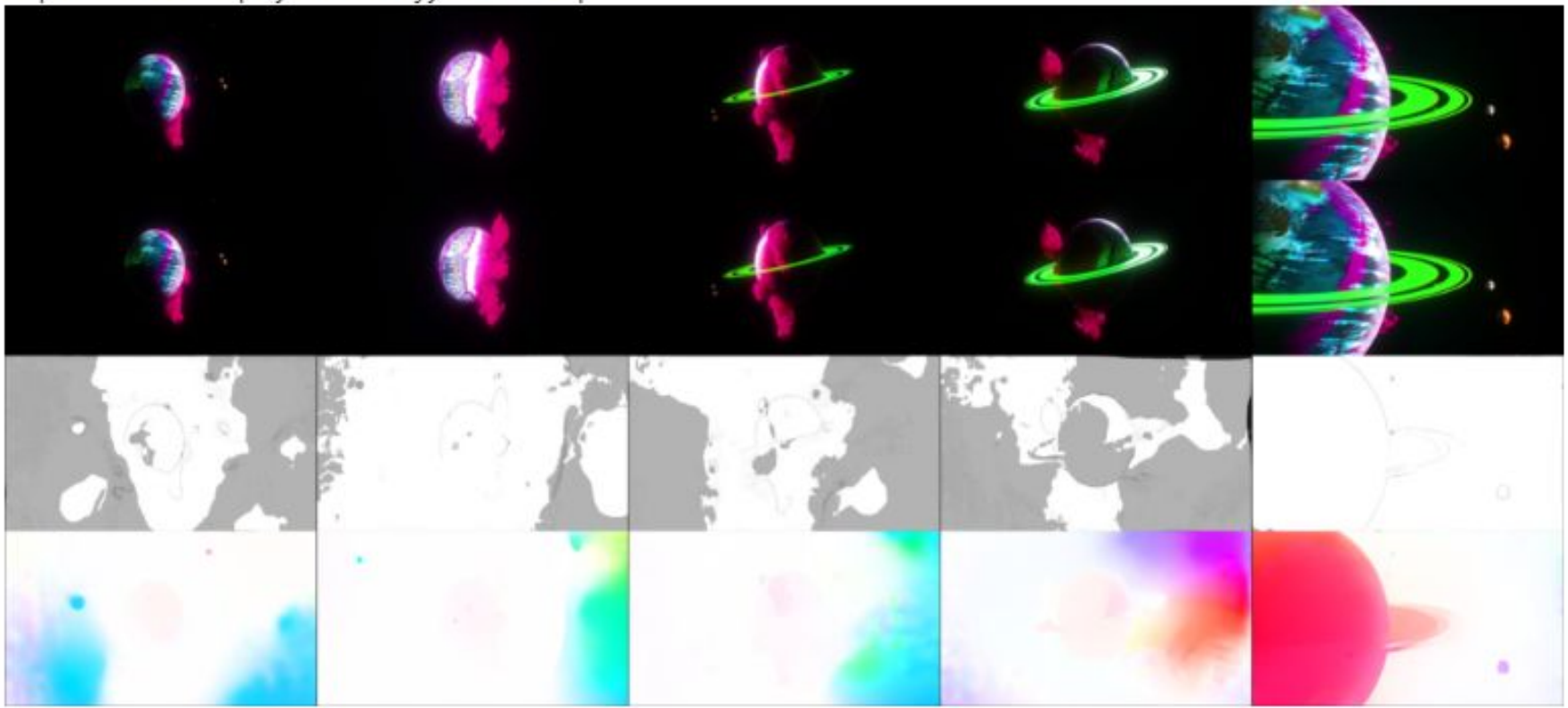

LORA & embedding paths
lora_dir: "/content/drive/MyDrive/models/loras
custom_embed_dir: "/content/drive/MyDrive/models/embeddings


stable_warpfusion.ipynb
File Edit View Insert Runtime Tools Help
Cannot save changes
+ Code + Text
Show code
Run all ⌘/Ctrl+F9
Run before ⌘/Ctrl+F8
Run the focused cell ⌘/Ctrl+Enter
Run selection ⌘/Ctrl+Shift+Enter

It takes ~30 minutes to start generating images. It depends how long the video is.

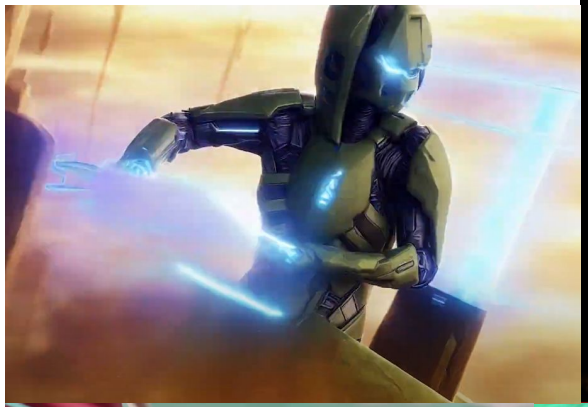


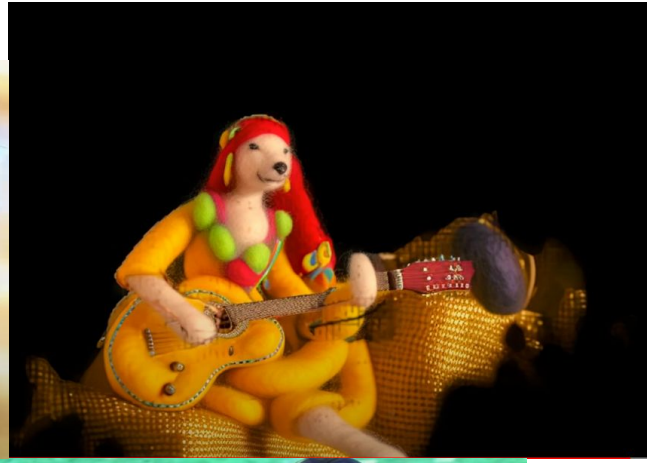
Samples from raw. alpha, consistency, and flow maps

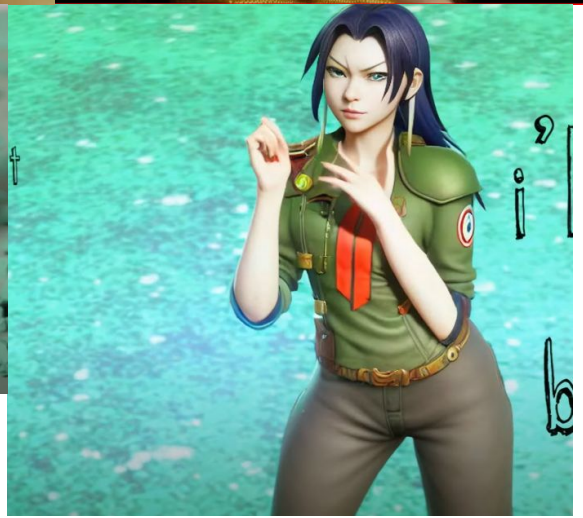


- I practiced by making fan art
- Eventually led to some paid work

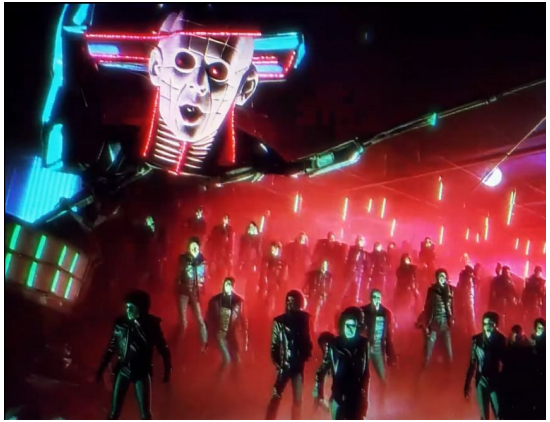
[Revolver Magazine](#)
[@g.m.zero](#)







[Look What I Did - Jekyll Island Fiat Scratch]



[Gromo - Prologue/The Vortex (Short Film)]

[Frost Children - HI 5]

[MIZMOR - No Place To Arrive]

Tools in my current workflow:

- Civitai.com (models, loras)
- Lexica.art
- Warpfusion
- Adobe Premiere Pro
- Adobe After Effects
- Topaz Video AI

Let me know if you have any questions.

Thanks!