

▼ MSE Regularized VQGAN + Clip (v.2)

by [@jbusted1](#)

Generates images from text prompts with VQGAN and CLIP (z+quantize method), with MSE regularization.

🎁 Improvements

- Added custom schedule of learning rates and mse weights to get more cohesive results
- Adjusted augmentation pipeline for better stability
- Add CLIP Prior technique to convert text prompts to image prompts for greater variety (highley recommended)
- Included several presets to help test out the technique, depending on your goals

Based on a notebook by Katherine Crowson ([@RiversHaveWings](#) who also trained the Clip prior, <https://github.com/crowsonkb>); with some UX help from [@remi_durant](#).

visitors 857

```
#@markdown Licensed under the MIT License           Licensed under the MIT License
# Copyright (c) 2022 Katherine Crowson and John David Pressman

# Permission is hereby granted, free of charge, to any person obtaining a copy
# of this software and associated documentation files (the "Software"), to deal
# in the Software without restriction, including without limitation the rights
# to use, copy, modify, merge, publish, distribute, sublicense, and/or sell
# copies of the Software, and to permit persons to whom the Software is
# furnished to do so, subject to the following conditions:

# The above copyright notice and this permission notice shall be included in
# all copies or substantial portions of the Software.

# THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
# IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
# FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
# AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
# LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,
# OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN
# THE SOFTWARE.
```

- ▶ First check what GPU you got and make sure it's a good one.

- Tier List: (K80 < T4 < P100 < V100 < A100)

[Show code](#)

Tesla T4, 15109 MiB, 15109 MiB

▼ Setup (Run Once)

▼ Download and install dependencies

installations

[Show code](#)

```
Cloning into 'CLIP'...
remote: Enumerating objects: 236, done.
remote: Total 236 (delta 0), reused 0 (delta 0), pack-reused 236
Receiving objects: 100% (236/236), 8.92 MiB | 25.16 MiB/s, done.
Resolving deltas: 100% (120/120), done.
Cloning into 'taming-transformers'...
remote: Enumerating objects: 1335, done.
remote: Total 1335 (delta 0), reused 0 (delta 0), pack-reused 1335
Receiving objects: 100% (1335/1335), 409.77 MiB | 50.13 MiB/s, done.
Resolving deltas: 100% (279/279), done.
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
Collecting ftfy
    Downloading ftfy-6.1.1-py3-none-any.whl (53 kB)
      [██████████] | 53 kB 1.5 MB/s
Requirement already satisfied: regex in /usr/local/lib/python3.7/dist-packages (2022.6.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (4.64.1)
Collecting omegaconf
    Downloading omegaconf-2.2.3-py3-none-any.whl (79 kB)
      [██████████] | 79 kB 4.5 MB/s
Collecting pytorch-lightning
    Downloading pytorch_lightning-1.7.7-py3-none-any.whl (708 kB)
      [██████████] | 708 kB 44.6 MB/s
Requirement already satisfied: wcwidth>=0.2.5 in /usr/local/lib/python3.7/dist-packages
Collecting antlr4-python3-runtime==4.9.*
    Downloading antlr4-python3-runtime-4.9.3.tar.gz (117 kB)
      [██████████] | 117 kB 71.3 MB/s
Requirement already satisfied: PyYAML>=5.1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: tensorboard>=2.9.1 in /usr/local/lib/python3.7/dist-packages
Collecting pyDeprecate>=0.3.1
    Downloading pyDeprecate-0.3.2-py3-none-any.whl (10 kB)
Requirement already satisfied: torch>=1.9.* in /usr/local/lib/python3.7/dist-packages (1.9.1)
Requirement already satisfied: typing-extensions>=4.0.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: fsspec[http]!=2021.06.0,>=2021.05.0 in /usr/local/lib/python3.7/dist-packages
Collecting torchmetrics>=0.7.0
    Downloading torchmetrics-0.10.0-py3-none-any.whl (529 kB)
      [██████████] | 529 kB 50.7 MB/s
Requirement already satisfied: packaging>=17.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: numpy>=1.17.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from torchmetrics>=0.7.0)
Requirement already satisfied: aiohttp!=4.0.0a0,!>4.0.0a1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: asynctest==0.13.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: charset-normalizer<3.0,>=2.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: grpcio>=1.24.3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: setuptools>=41.0.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in /usr/local/lib/python3.7/dist-packages
```

```

Requirement already satisfied: protobuf<3.20,>=3.9.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: wheel>=0.26 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: werkzeug>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: absl-py>=0.4 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: cachetools<5.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: importlib-metadata>=4.4 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: urllib3!=1.25.0,!>=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from
Building wheels for collected packages: antlr4-python3-runtime
  Building wheel for antlr4-python3-runtime (setup.py) ... done
  Created wheel for antlr4-python3-runtime: filename=antlr4_python3_runtime-4.9.3-py3-none-any.whl
  Stored in directory: /root/.cache/pip/wheels/8b/8d/53/2af8772d9aec614e3fc65e53d4a993ac
Successfully built antlr4-python3-runtime
Installing collected packages: torchmetrics, pyDeprecate, antlr4-python3-runtime, pytorch
Successfully installed antlr4-python3-runtime-4.9.3 ftfy-6.1.1 omegaconf-2.2.3 pyDeprecate-0.0.1
WARNING: The following packages were previously imported in this runtime:
  [pydevd_plugins]
You must restart the runtime in order to use newly installed versions.

```

RESTART RUNTIME

```

Cloning into 'v-diffusion-pytorch'...
remote: Enumerating objects: 155, done.
remote: Counting objects: 100% (155/155)   ----

```

model downloads

[Show code](#)

% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current				
			Dload	Upload	Total	Spent	Left	Speed			
100	331M	100	331M	0	0	37.6M	0	0:00:08	0:00:08	--::--	40.9M
% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current				
			Dload	Upload	Total	Spent	Left	Speed			
0	0	0	0	0	0	0	--::--	--::--	--::--	--::--	0
100	692	100	692	0	0	909	0	--::--	--::--	--::--	909
% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current				
			Dload	Upload	Total	Spent	Left	Speed			
0	0	0	0	0	0	0	--::--	--::--	--::--	--::--	0
100	934M	100	934M	0	0	14.8M	0	0:01:02	0:01:02	--::--	15.9M

Running output: python -m torch.distributed.launch --nproc_per_node=4 --master_port=29501 train.py

Utilities

Downloading pytorch_optimizer-1.1.2-py3-none-any.whl (56 kB)

update path

[Show code](#)

```
!pip install -q git+https://github.com/CompVis/taming-transformers@main --no-cache-dir --force-reinstall
```

imports

[Show code](#)

vqgan related

[Show code](#)

prior related

[Show code](#)

random choice augment module

[Show code](#)

remi utility

[Show code](#)

▼ Make Some Art!

▼ Preset Definitions

```
#@title  Preset Definitions
#@markdown This cell must be run once. If you want to edit the presets, you can expand this cell and edit the code directly.

presets = {
    "fun":argparse.Namespace(
        init_weight = 2.0,
        min_weight = 0.00,
        max_step = 700,

        step_size = 0.45,
        min_step_size = 0.45,

        # cutouts / crops scheme
        cutn = 64,
        cut_pow = 1,
```

This cell must be run once. If you want to edit the presets, you can expand this cell and edit the code directly.

```
# mse settings
change_orgin_at = 26,
mse_withzeros = True,
mse_decay_rate = 50,
mse_epoches = 10,
mse_quantize = False,

use_prior = False,
custom_cropping_scheme = False,
),

"high quality":argparse.Namespace(
    init_weight = 2.0,
    min_weight = 0.00,
    max_step = 700,

    step_size = 0.95,
    min_step_size = 0.25,

    # cutouts / crops scheme
    cutn = 64,
    cut_pow = 1,

    # mse settings
    change_orgin_at = 460,
    mse_withzeros = True,
    mse_decay_rate = 50,
    mse_epoches = 10,
    mse_quantize = False,

    use_prior = False,
    custom_cropping_scheme = False,
),
"fast":argparse.Namespace(
    init_weight = 1.7,
    min_weight = 0.00,
    max_step = 400,

    step_size = 0.95,

    # cutouts / crops scheme
    cutn = 32,
    cut_pow = 1,

    # mse settings
    change_orgin_at = 460,
    mse_withzeros = True,
    mse_decay_rate = 25,
    mse_epoches = 10,
    mse_quantize = False,
```

```
        use_prior = False,  
        custom_cropping_scheme = True,  
    )  
}
```

Do the Run

What do you want to see?

Enter your prompts as an array of strings to help guide VQGAN + CLIP.

- if you want to try using CLIP prior, enable `use_prior`. This will use a diffusion prior to translate your text prompts into image prompts which should result in greater diversity of outputs.

prompts: "I find myself on a beach, with many beachfront mansions nearby. My friends and I dec

use_prior:

seed: -1

Display dimensions need to be divisible by 16

width: 640

height: 512

Do you want to use an init image instead of a randomized start?

init_image: "Insert text here"

How frequently do you want to see the outputs?

display_freq: 25

Are you going to want to make a video after of the full generation?

record_generation:

Presets

Choose a preset to override some of the settings below with premade configurations, or leave as none for complete control.

- **high quality** - take it slow and steady, good results by iteration 600
- **fun** - forces the generation to commit to early results. color is more vivid throughout the generation, might work better for some prompts
- **fast** - same as high quality but faster, might be a little messy, full results by iteration 300

`use_preset: fun`

Settings

Note: If you have selected a preset above, it may override some of these settings.

Learning Rates

The `step_size` is used as the initial learning rate, and at each mse epoch will decay until it reaches the `min_step_size`. Higher step sizes may make the generation less stable, but this is a speed/stability tradeoff.

`step_size: 0.95`

`min_step_size: 0.25`

Cuts

Enable `custom_cropping_scheme` to perform more evenly distributed crops across the image (bias decided via `cut_pow`), or uncheck to focus on details near the center of the image (using mostly large and very small crops).

`custom_cropping_scheme:`

`cut_pow: 1.0`

Number of cuts to make, you can reduce this to save memory but it may make the generation take longer.

`cutn: 64`

MSE Schedule

The mse weight determines how strongly the generator pays attention to the mse loss. It starts at `init_weight` and then will decay at each mse epoch until it reaches `min_weight`.

`init_weight: 2.0`

`min_weight: 0.0`

lower values fills the rest of the details better, but restricts significant changes after the numbered step

`change_origin_at: 460`

whether or not quantize the reference image before changing it

`mse_quantize:`

This gives finer control on how the mse weight decays and changes during the generation.

`mse_decay_rate: 50`

`mse_epochs: 10`

`max_step: 701`

Other MSE options:

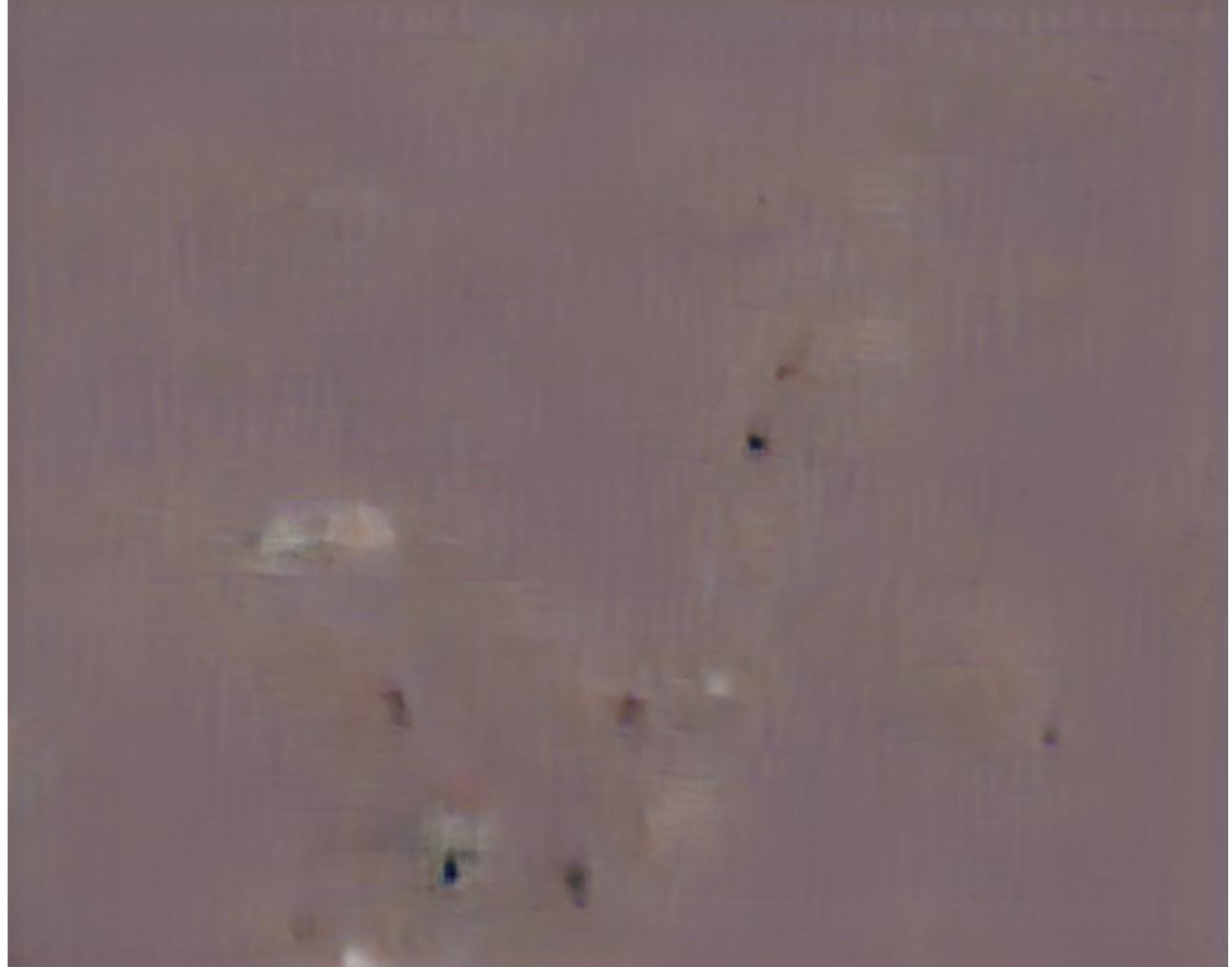
`mse_withzeros:`

[Show code](#)

```
Using device: cuda:0
using prompts: ['I find myself on a beach, with many beachfront mansions nearby. My fri
Working with z of shape (1, 256, 16, 16) = 65536 dimensions.
/usr/local/lib/python3.7/dist-packages/torchvision/models/_utils.py:209: UserWarning: Tl
  f"The parameter '{pretrained_param}' is deprecated since 0.13 and will be removed in <
/usr/local/lib/python3.7/dist-packages/torchvision/models/_utils.py:223: UserWarning: Ar
  warnings.warn(msg)
Downloading: "https://download.pytorch.org/models/vgg16-397923af.pth" to /root/.cache/tc
100%                                         528M/528M [00:05<00:00, 96.3MB/s]

Downloading vgg_lpipps model from https://heibox.uni-heidelberg.de/f/607503859c864bc1b30/
8.19kB [00:00, 82.9kB/s]
loaded pretrained LPIPS loss from taming/modules/autoencoder/lpipps/vgg.pth
VQLPIPSWithDiscriminator running with hinge loss.
Restored from vqgan_imagenet_f16_16384.ckpt
100%|██████████| 338M/338M [00:02<00:00, 165MiB/s]
Using seed 86622 for generation
700/? [18:11<00:00, 1.53s/it]

i: 0, loss: 0.978618, losses: 0.042, 0.936618
averaged:
```



```
i: 25, loss: 1.10199, losses: 0.25246, 0.849525
averaged:
```





```
replace z_org  
i: 50, loss: 1.05351, losses: 0.221181, 0.832324  
averaged:
```

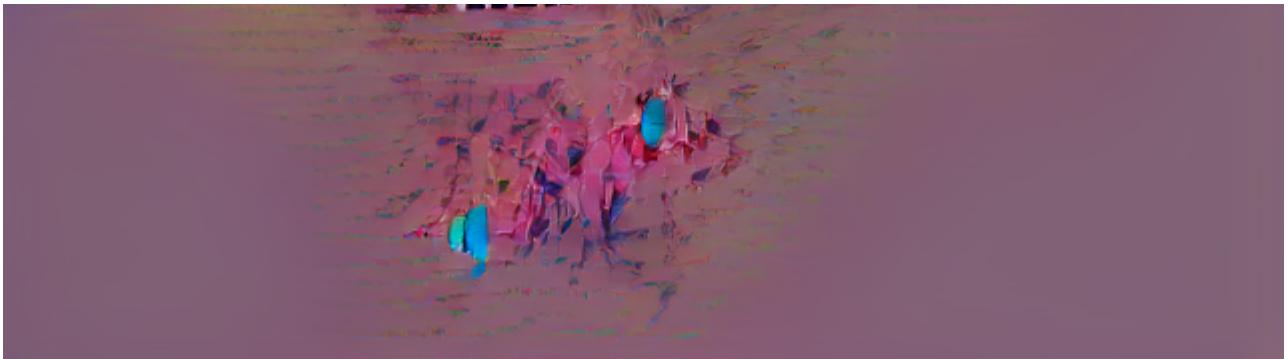


```
restarting optimization 50 and learning rate 0.45
i: 75, loss: 0.976007, losses: 0.170325, 0.805681
averaged:
```



```
replace z_org
i: 100, loss: 0.988998, losses: 0.187382, 0.801616
averaged:
```

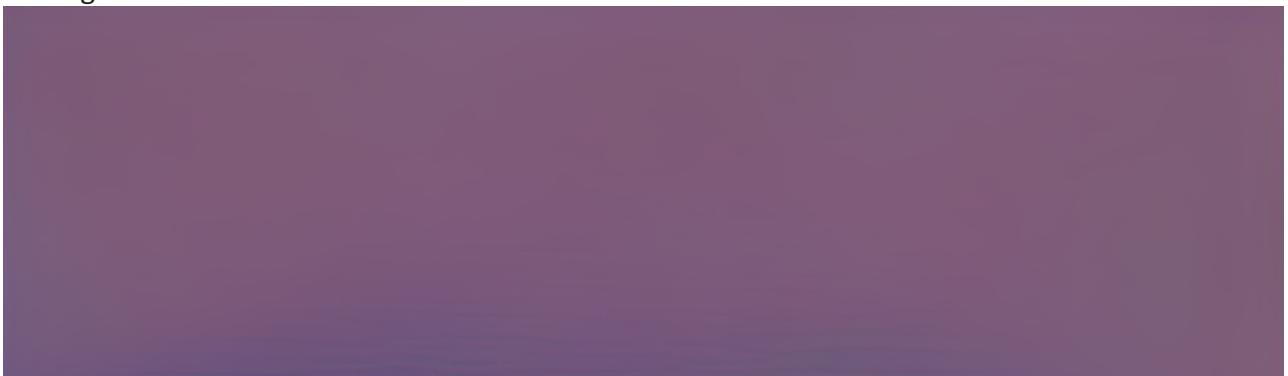




```
restarting optimization 100 and learning rate 0.45
i: 125, loss: 0.938787, losses: 0.149768, 0.789019
averaged:
```



```
replace z_org
i: 150, loss: 0.930677, losses: 0.123285, 0.807391
averaged:
```





restarting optimization 150 and learning rate 0.45

i: 175, loss: 0.9169, losses: 0.114824, 0.802076

averaged:



replace z_org

i: 200, loss: 0.880872, losses: 0.0954771, 0.785395

averaged:



restarting optimization 200 and learning rate 0.45

i: 225, loss: 0.882312, losses: 0.103983, 0.778329

averaged:





```
replace z_org  
i: 250, loss: 0.862777, losses: 0.0847561, 0.778021  
averaged:
```



```
restarting optimization 250 and learning rate 0.45  
i: 275, loss: 0.873195, losses: 0.0884883, 0.784707  
averaged:
```





replace z_org

i: 300, loss: 0.845262, losses: 0.0727392, 0.772522

averaged:



restarting optimization 300 and learning rate 0.45

i: 325, loss: 0.8765, losses: 0.102995, 0.773505

averaged:





replace z_org

i: 350, loss: 0.840979, losses: 0.0740056, 0.766974

averaged:



restarting optimization 350 and learning rate 0.45

i: 375, loss: 0.867594, losses: 0.0994794, 0.768115

averaged:



```
replace z_org  
i: 400, loss: 0.829093, losses: 0.0621215, 0.766971  
averaged:
```





restarting optimization 400 and learning rate 0.45
i: 425, loss: 0.867035, losses: 0.104365, 0.76267
averaged:



replace z_org
i: 450, loss: 0.809806, losses: 0.0497762, 0.760029
averaged:





restarting optimization 450 and learning rate 0.45

i: 475, loss: 0.882881, losses: 0.122406, 0.760475

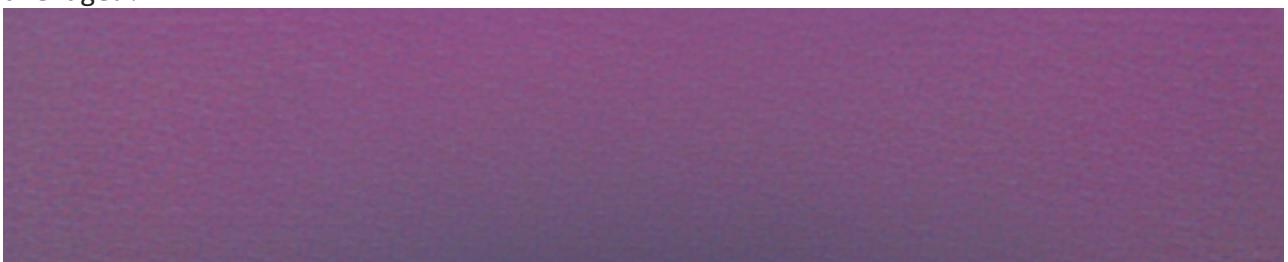
averaged:



replace z_org

i: 500, loss: 0.764987, losses: 0, 0.764987

averaged:





i: 525, loss: 0.762047, losses: 0, 0.762047
averaged:



i: 550, loss: 0.753432, losses: 0, 0.753432
averaged:



i: 575, loss: 0.75052, losses: 0, 0.75052
averaged:





i: 600, loss: 0.744476, losses: 0, 0.744476
averaged:



i: 625, loss: 0.743711, losses: 0, 0.743711
averaged:

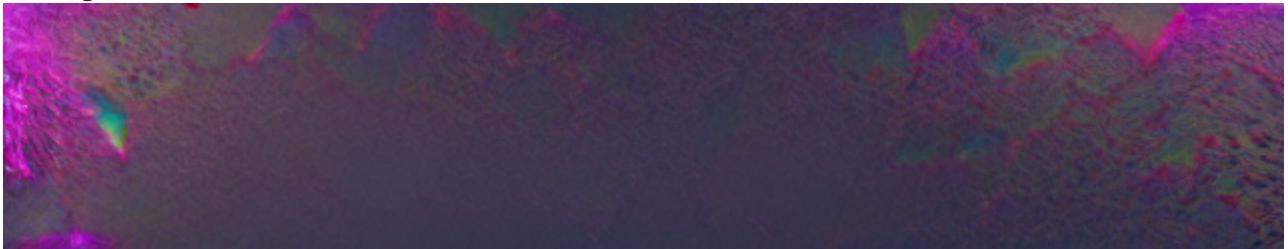




i: 650, loss: 0.737593, losses: 0, 0.737593
averaged:



i: 675, loss: 0.737103, losses: 0, 0.737103
averaged:



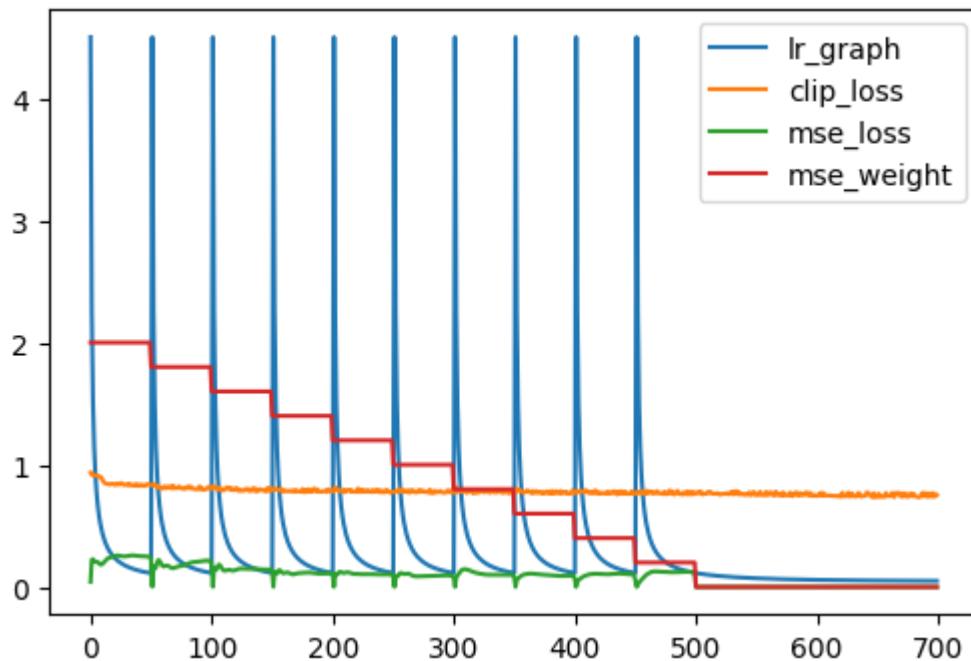
▼ Inspect Losses

Show Graph

This graph will show the learning rate, clip loss, and mse weight during the generation.

[Show code](#)

0.49013056167154767



🔍 Show Graph

This graph will show the learning rate, clip loss, and mse weight during the generation.

[Show code](#)

```
<matplotlib.legend.Legend at 0x7fb3f0170f10>
```

▼ Render Video

Make a video of your generation. For this to work, `record_generation` must be enabled.



▼ Save Video Frames to Disk

```
# @title  Save Video Frames to Disk
if args.record_generation:
    for vid_index, out_ in enumerate(tqdm(video_frames)):
        TF.to_pil_image(out_).save(f'/content/vids/{vid_index}.png')
```

100% 700/700 [01:39<00:00, 8.44it/s]

0 100 200 300 400 500 600 700

▶ Render and Download

`video: "/content/movie1024.mp4"`

`framerate: 24`

`sizexy: "640x512"`

[Show code](#)

```
/content/vids
ffmpeg version 3.4.11-0ubuntu0.1 Copyright (c) 2000-2022 the FFmpeg developers
  built with gcc 7 (Ubuntu 7.5.0-3ubuntu1~18.04)
configuration: --prefix=/usr --extra-version=0ubuntu0.1 --toolchain=hardened --libdir=
libavutil      55. 78.100 / 55. 78.100
libavcodec     57.107.100 / 57.107.100
libavformat    57. 83.100 / 57. 83.100
libavdevice    57. 10.100 / 57. 10.100
libavfilter     6.107.100 / 6.107.100
libavresample   3.  7.  0 / 3.  7.  0
libswscale      4.  8.100 / 4.  8.100
libswresample   2.  9.100 / 2.  9.100
libpostproc    54.  7.100 / 54.  7.100
Input #0, image2, from '%d.png':
  Duration: 00:00:28.00, start: 0.000000, bitrate: N/A
    Stream #0:0: Video: png, rgb24(pc), 640x512, 25 fps, 25 tbr, 25 tbn, 25 tbc
Stream mapping:
  Stream #0:0 -> #0:0 (png (native) -> h264 (libx264))
Press [q] to stop, [?] for help
[libx264 @ 0x55844d067e00] using cpu capabilities: MMX2 SSE2Fast SSSE3 SSE4.2 AVX FMA3 F
[libx264 @ 0x55844d067e00] profile High, level 3.0
[libx264 @ 0x55844d067e00] 264 - core 152 r2854 e9a5903 - H.264/MPEG-4 AVC codec - Copy]
Output #0, mp4, to '/content/movie1024.mp4':
  Metadata:
    encoder : Lavf57.83.100
  Stream #0:0: Video: h264 (libx264) (avc1 / 0x31637661), yuv420p, 640x512, q=-1--1, 2
  Metadata:
    encoder : Lavc57.107.100 libx264
  Side data:
    cpb: bitrate max/min/avg: 0/0/0 buffer size: 0 vbv_delay: -1
frame= 700 fps= 35 q=-1.0 Lsize= 6807kB time=00:00:29.04 bitrate=1920.1kbits/s speed
video:6798kB audio:0kB subtitle:0kB other streams:0kB global headers:0kB muxing overhead
[libx264 @ 0x55844d067e00] frame I:13 Avg QP:17.48 size: 28714
[libx264 @ 0x55844d067e00] frame P:19 Avg QP:19.50 size: 22019
```

▼ Delete All Frames from Folder

```
#@title    Delete All Frames from Folder
delete_images = False# @param{type:"boolean"}
```

delete_images: 

```
if delete_images:
  %cd vids
  %rm *.png
  %cd ..
```

```
[libx264 @ 0x55844d067e00] Weighted P-Frames: Y:31.5% UV:25.4%
[libx264 @ 0x55844d067e00] ref P L0: 50.7% 25.5% 15.5% 6.8% 1.5%
[libx264 @ 0x55844d067e00] ref B L0: 90.4% 7.8% 1.8%
[libx264 @ 0x55844d067e00] ref B L1: 95.4% 4.6%
[libx264 @ 0x55844d067e00] kb/s:1909.13
/content
```