EncDiff Replication: Debug

1. Retrained model link mistake: led to the dataset link by mistake → Train from scratch
   a. file name misalign: main.py —> main_val.py; dis.py → disdata.py
   b. Said dataset 3dshape.h5 but main_val.py uses 3dshape.npz
2. Modules not uploaded → Commented
   a. No module named 'ldm.modules.diffusionmodules.vct_encoder'
3. Current env has conflict with H100（PyTorch 1.7.0/CUDA 11.0) → New Env
   a. Update `env-H100.yaml`, and `/configs/latent-diffusion/shapes3d-vq-4-16-encdiff.yaml`,and `/configs/autoencoder/shapes3d_vq_4_16.yaml`, …
   b. Update `main_val.py`, and `/src/taming-transformers/main.py`, and `ldm/models/diffusion/ddpm_enc.py` and `/ldm/data/disdata.py`, … modify some functions that have version conflict in packaging/transformers/lightning…
4. Training settings
   a. Num_workers from 256 → 8 since we use H100
   b. Add max_epoch.
   c. Batch size 128 (paper: 64; but github repo default 128)
   d. Visualization & Checkpoints path: `/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-04T09-59-07_shapes3d-vq-4-16-encdiff23`

Datasets

1. Shapes3D
    a. 480,000 RGB images (64 x 64) of 3D objects in a room with perfectly controlled generative factors.
    b. Factors: 6 ground-truth factors (floor color, wall color, object color, scale, shape, orientation).
2. Cars3D
    a. 17,568 RGB images (64 x 64) of CAD models of cars rendered from multiple viewpoints.
    b. Factors: 3 ground-truth factors (elevation, azimuth, object ID).
3. MPI3D (Toy)
    a. 1,032,192 RGB images (64 x 64) featuring a robotic arm holding various objects in a simplified environment.
    b. Factors: 7 ground-truth factors (object color, shape, size, camera height, background color, two horizontal/vertical rotation axes).

Performance: **Shape3D > MPI3D > Cars3D**

- Cars3D challange: Complex non-primitive geometries and highly unbalanced factor distributions across 183 car models
- MPI3D challenge: Massive scale combined with subtle mechanical variations of the robotic arm requires intense computational resources for convergence

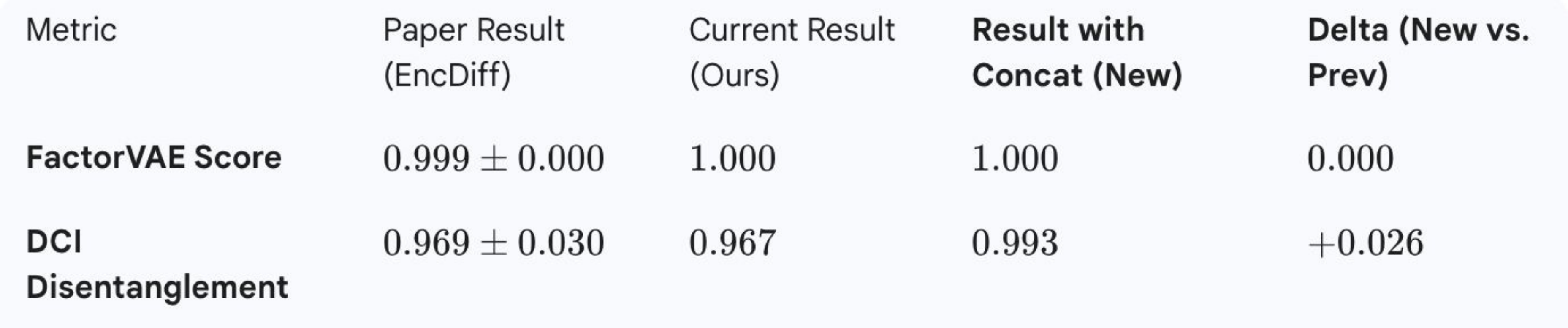
Add disentangled repr concat

1. modify ldm/models/autoencoder.py
2. modify ldm/models/diffusion/ddpm_enc.py
3. modify 3 config files

New Shape3D: Performance

Path:

```
/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-10T07-42-42_shapes3d-vq-4-16-encdiff23/metrics_sin/33750.json
```

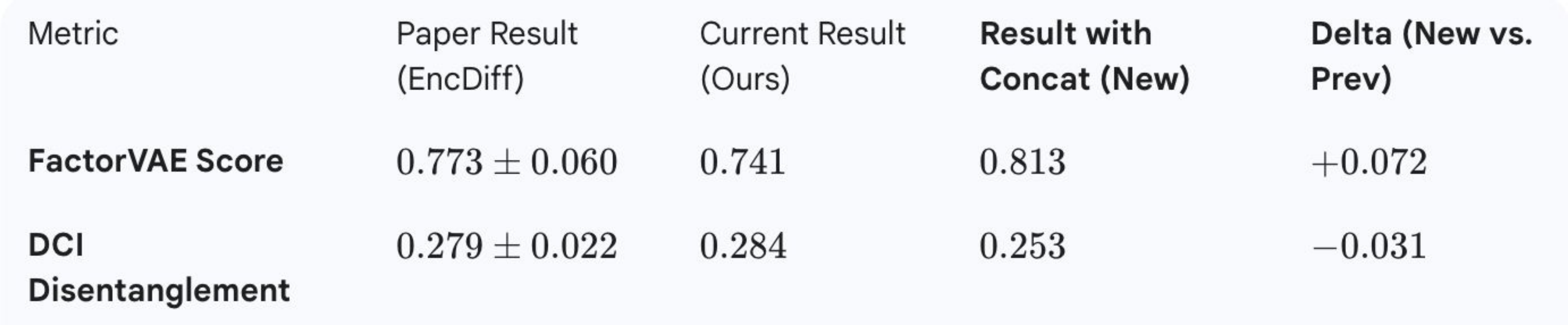


Metric	Paper Result (EncDiff)	Current Result (Ours)	**Result with Concat (New)**	**Delta (New vs. Prev)**
FactorVAE Score	0.999 ± 0.000	1.000	1.000	0.000
DCI Disentanglement	0.969 ± 0.030	0.967	0.993	$+0.026$

New Cars3D: Performance

Path:

`/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-10T08-25-57_cars3d-vq-4-16-encdiff23/metrics_sin/20595.json`

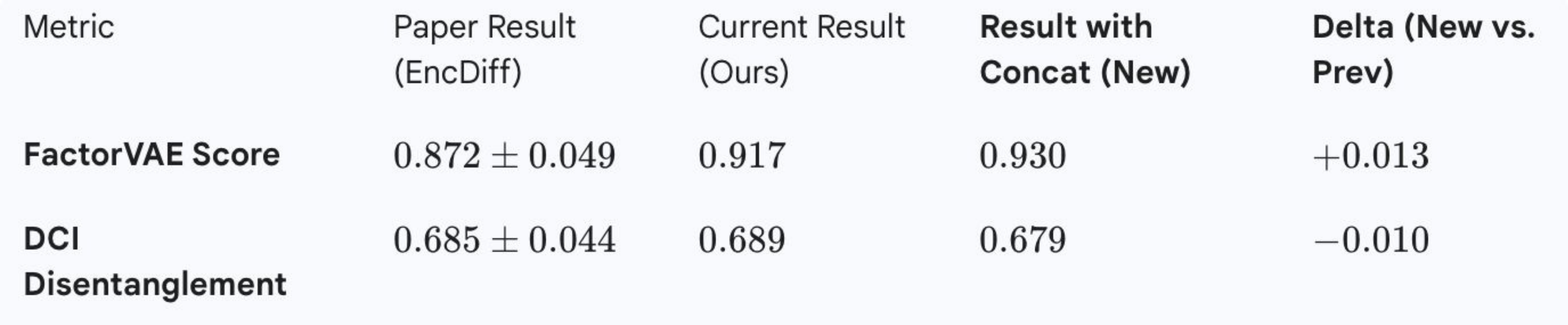


Metric	Paper Result (EncDiff)	Current Result (Ours)	**Result with Concat (New)**	**Delta (New vs. Prev)**
FactorVAE Score	0.773 ± 0.060	0.741	0.813	$+0.072$
DCI Disentanglement	0.279 ± 0.022	0.284	0.253	-0.031

New MPI3D-Toy: Performance

Path:

```
/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-05T06-14-44_mpi3d-vq-4-16-encdiff23/metrics_sin/81000.json
```

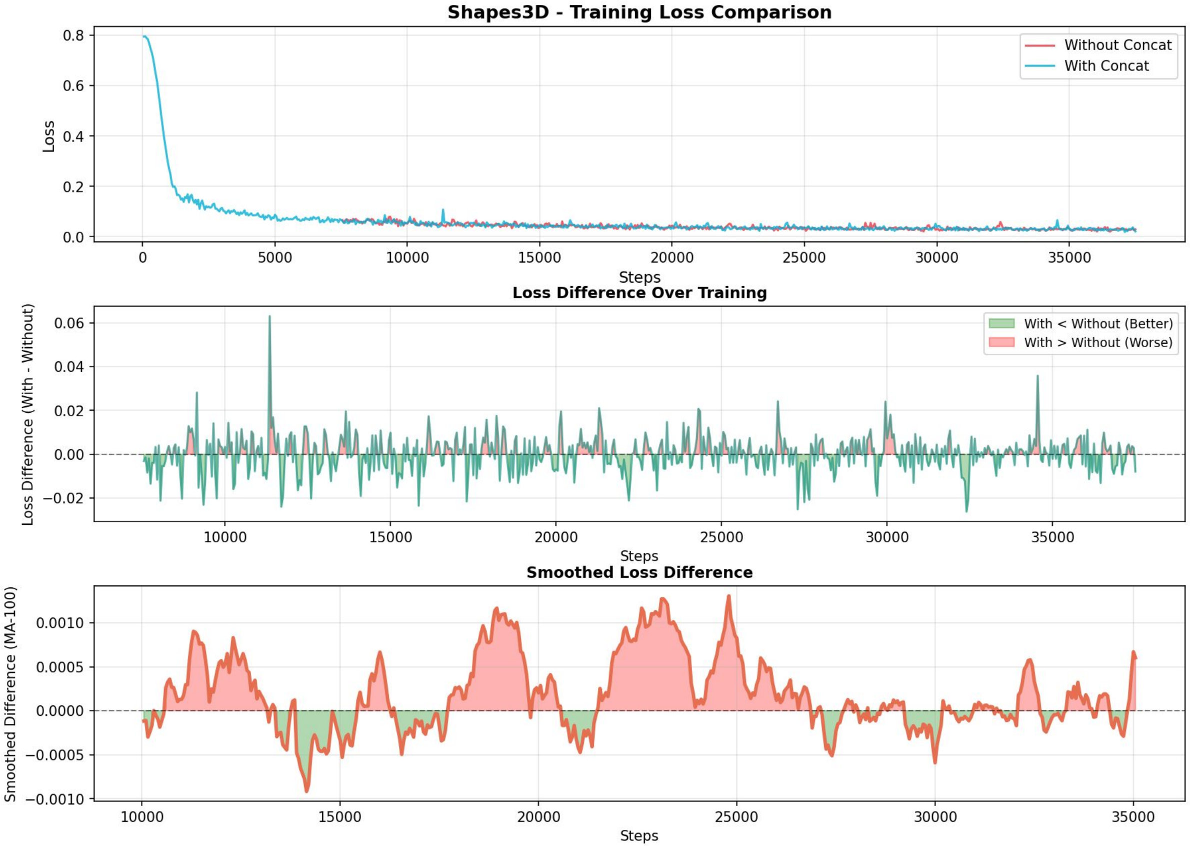


Metric	Paper Result (EncDiff)	Current Result (Ours)	**Result with Concat (New)**	**Delta (New vs. Prev)**
FactorVAE Score	0.872 ± 0.049	0.917	0.930	$+0.013$
DCI Disentanglement	0.685 ± 0.044	0.689	0.679	-0.010

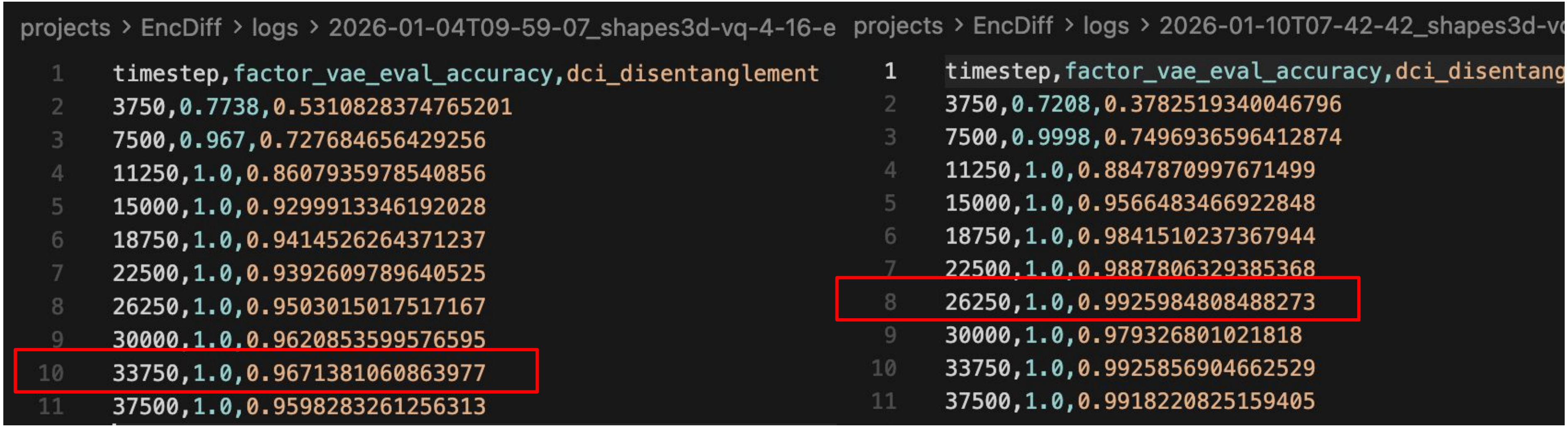
Explanation on following swapping images

First row: SRC(original images 0-7)
First Col: TRT(target image 0)
Second-last row: swap one factor from 20 latent factors

Shape3D: Training Loss Comparison



Shape3D: Results from each timesteps



projects › EncDiff › logs › 2026-01-04T09-59-07_shapes3d-vq-4-16-e

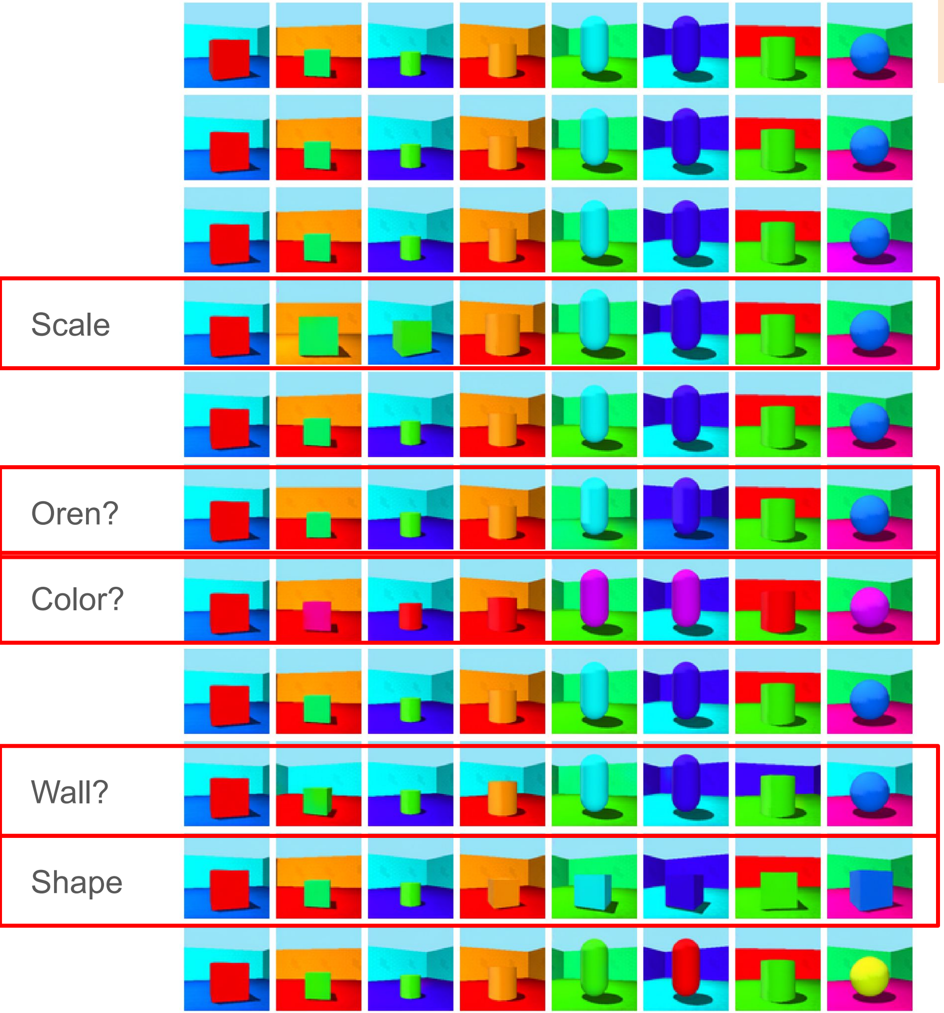
```
timestep,factor_vae_eval_accuracy,dci_disentanglement
3750,0.7738,0.5310828374765201
7500,0.967,0.727684656429256
11250,1.0,0.8607935978540856
15000,1.0,0.9299913346192028
18750,1.0,0.9414526264371237
22500,1.0,0.9392609789640525
26250,1.0,0.9503015017517167
30000,1.0,0.9620853599576595
33750,1.0,0.9671381060863977
37500,1.0,0.9598283261256313
```

projects › EncDiff › logs › 2026-01-10T07-42-42_shapes3d-vc

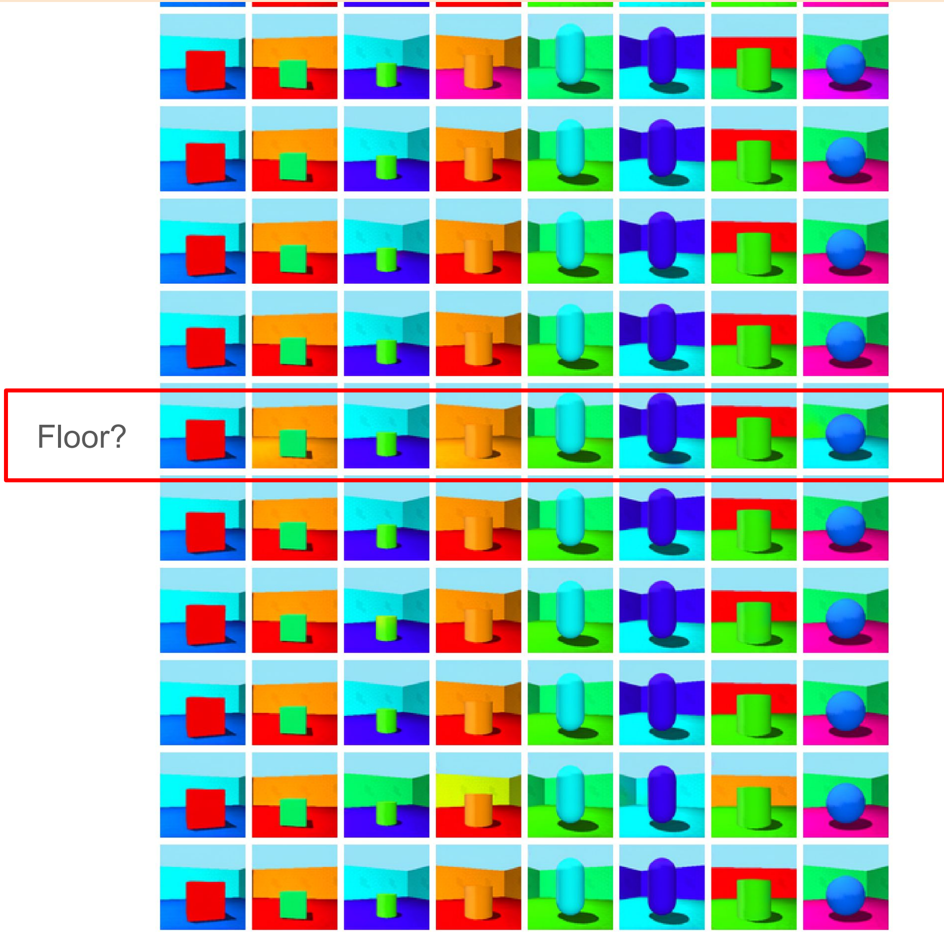
```
timestep,factor_vae_eval_accuracy,dci_disentang
3750,0.7208,0.3782519340046796
7500,0.9998,0.7496936596412874
11250,1.0,0.8847870997671499
15000,1.0,0.9566483466922848
18750,1.0,0.9841510237367944
22500,1.0,0.9887806329385368
26250,1.0,0.9925984808488273
30000,1.0,0.979326801021818
33750,1.0,0.9925856904662529
37500,1.0,0.9918220825159405
```

Baseline

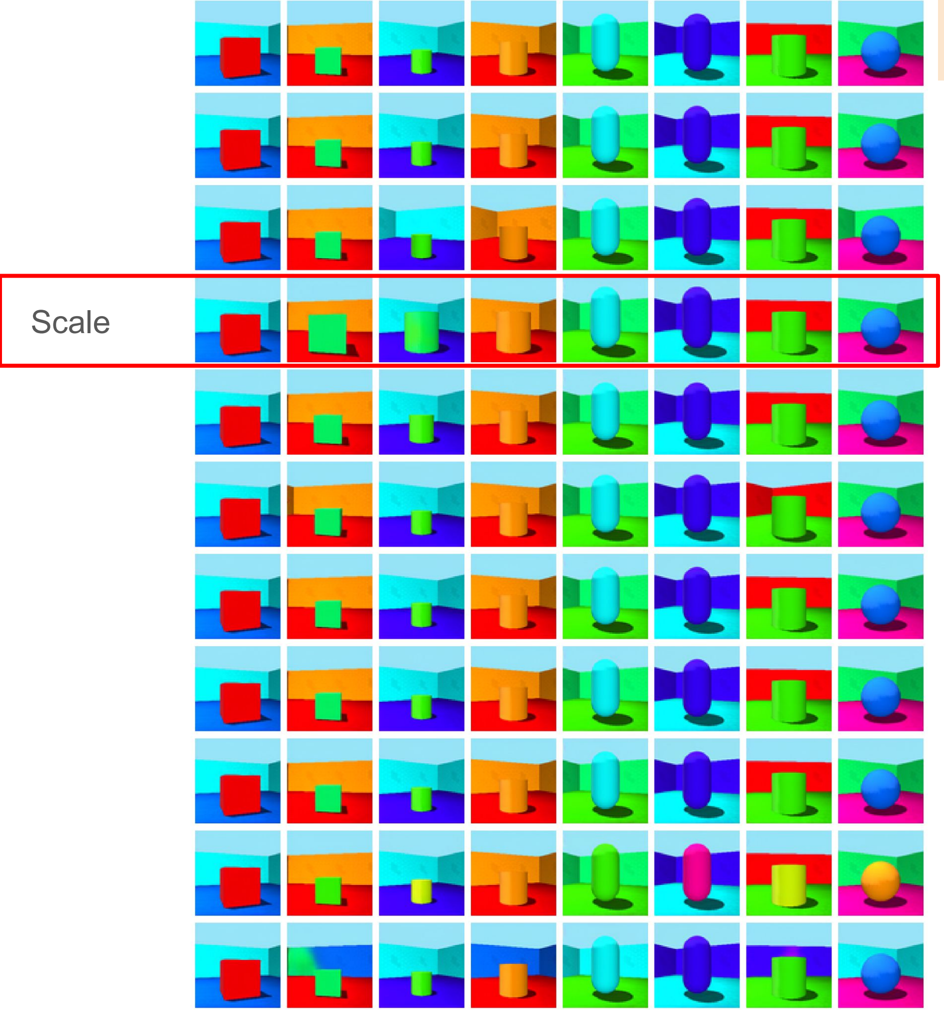
New

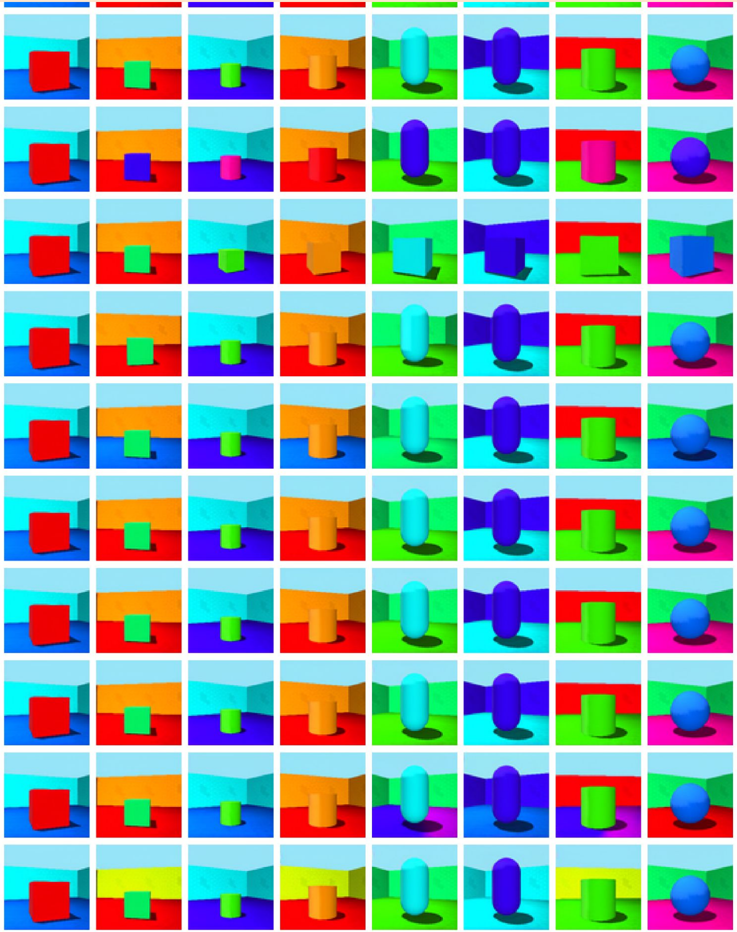


/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-04T09-59-07_shapes3d-vq-4-16-encdiff23/swap_images_last
37500 timesteps, baseline for shape3D

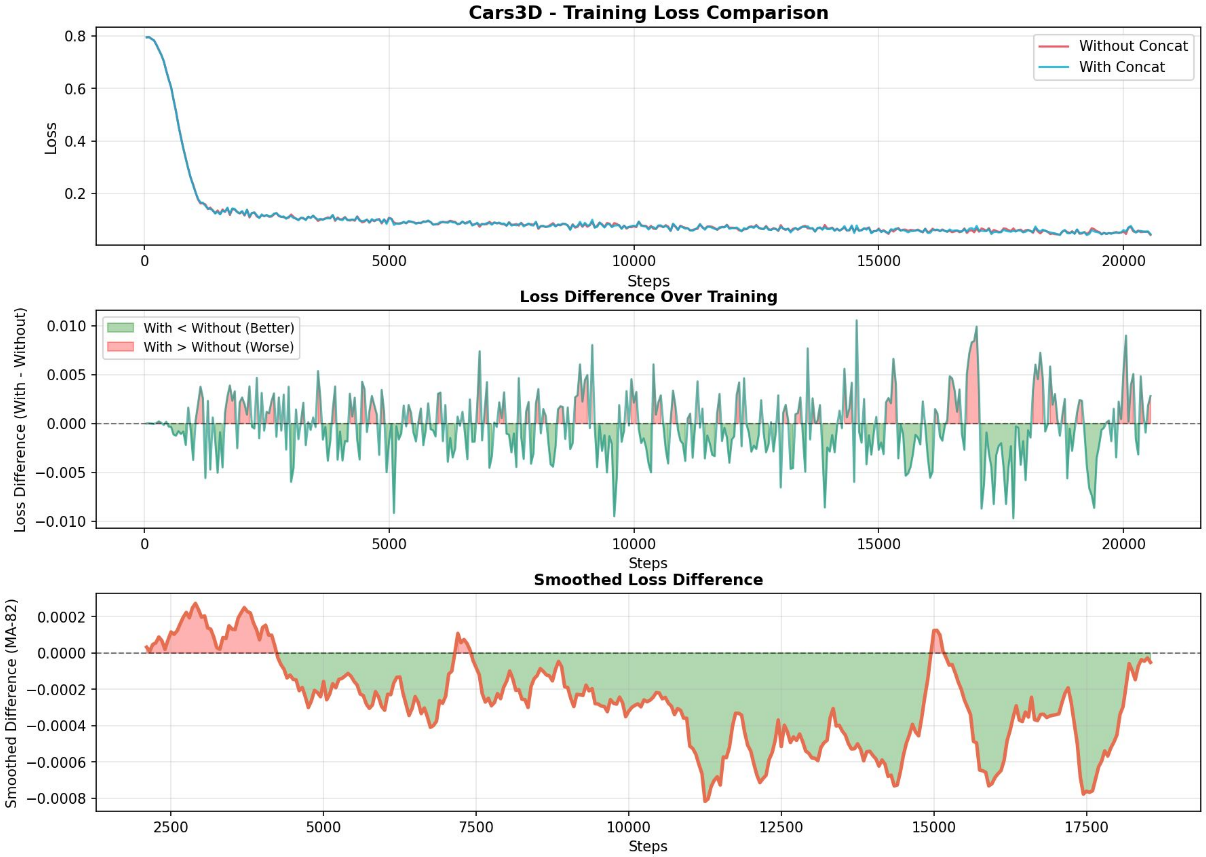


/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-10T07-42-42_shapes3d-vq-4-16-encdiff23/swap_images_last
37500 timesteps, w/ concat for shape3D

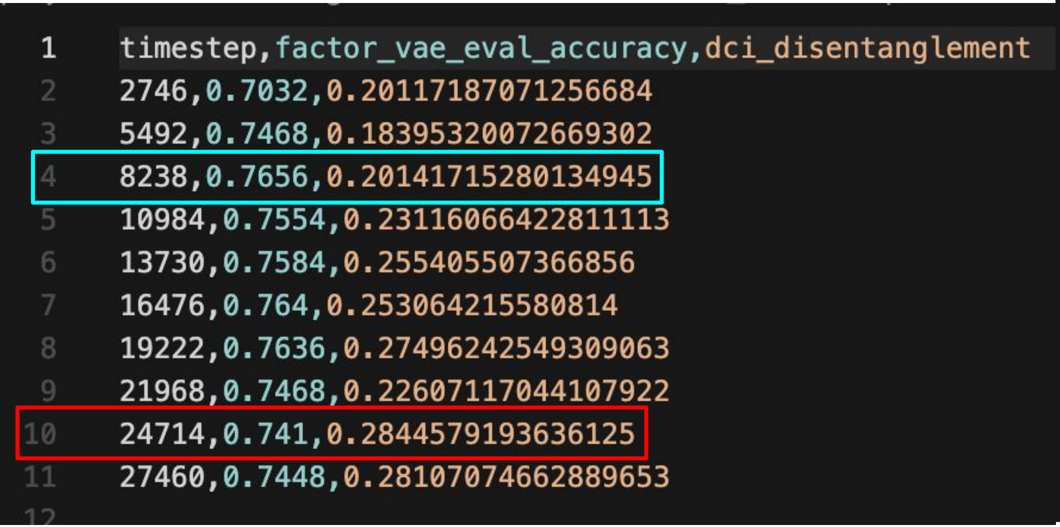




Cars3D: Training Loss Comparison

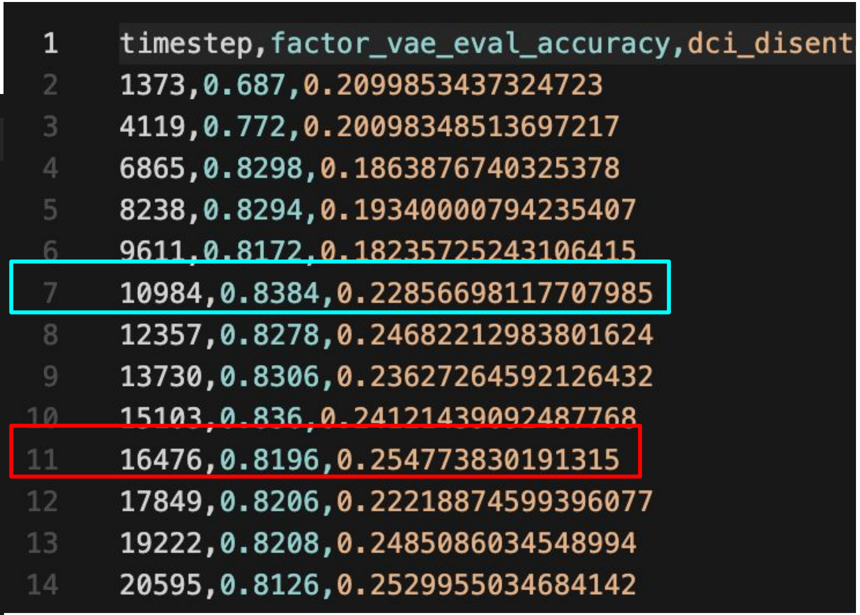


Cars3D: Results from each timesteps



```
timestep,factor_vae_eval_accuracy,dci_disentanglement
2746,0.7032,0.20117187071256684
5492,0.7468,0.18395320072669302
8238,0.7656,0.20141715280134945
10984,0.7554,0.23116066422811113
13730,0.7584,0.255405507366856
16476,0.764,0.253064215580814
19222,0.7636,0.27496242549309063
21968,0.7468,0.22607117044107922
24714,0.741,0.2844579193636125
27460,0.7448,0.28107074662889653
```

Baseline

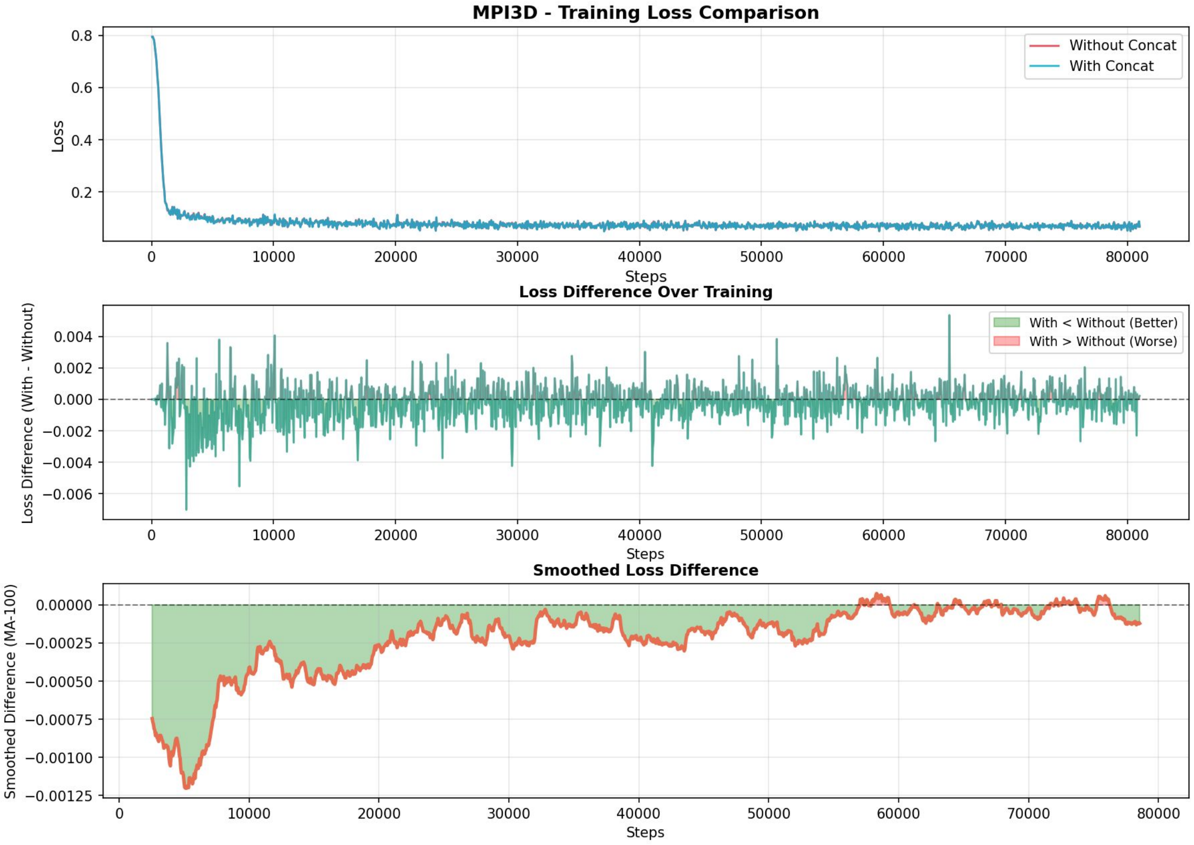


```
timestep,factor_vae_eval_accuracy,dci_disent
1373,0.687,0.2099853437324723
4119,0.772,0.20098348513697217
6865,0.8298,0.1863876740325378
8238,0.8294,0.19340000794235407
9611,0.8172,0.18235725243106415
10984,0.8384,0.22856698117707985
12357,0.8278,0.24682212983801624
13730,0.8306,0.23627264592126432
15103,0.836,0.24121439092487768
16476,0.8196,0.254773830191315
17849,0.8206,0.22218874599396077
19222,0.8208,0.2485086034548994
20595,0.8126,0.2529955034684142
```

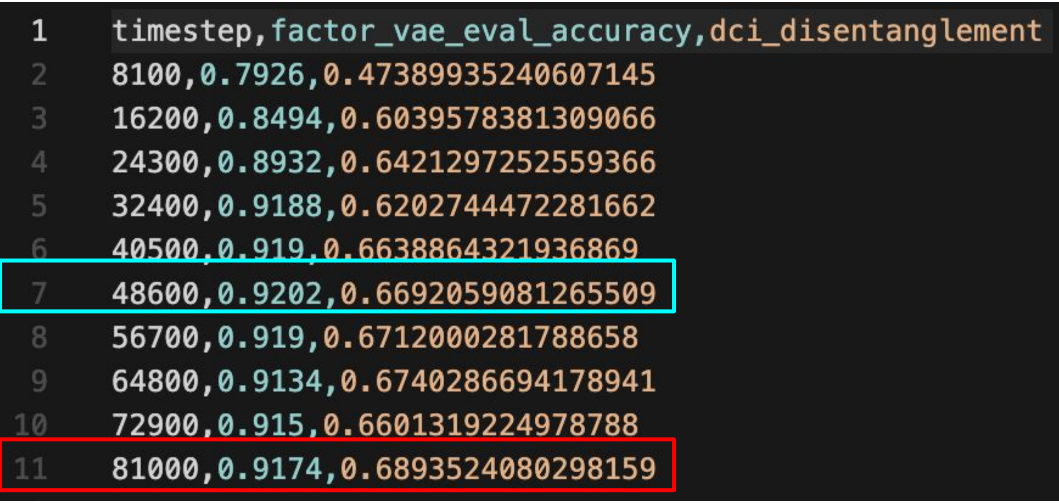
New

* Top 10 for either FactorVAE or DCI

MPI3D: Training Loss Comparison

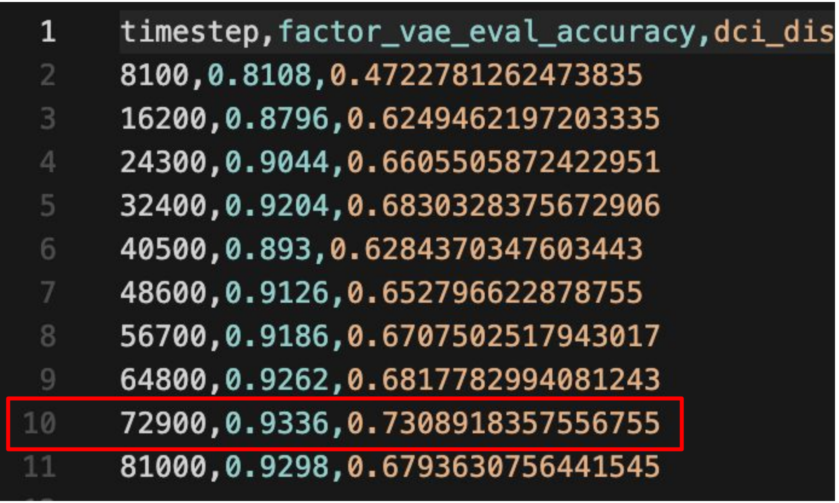


MPI3D: Results from each timesteps



```
timestep,factor_vae_eval_accuracy,dci_disentanglement
8100,0.7926,0.47389935240607145
16200,0.8494,0.6039578381309066
24300,0.8932,0.6421297252559366
32400,0.9188,0.6202744472281662
40500,0.919,0.6638864321936869
48600,0.9202,0.6692059081265509
56700,0.919,0.6712000281788658
64800,0.9134,0.6740286694178941
72900,0.915,0.6601319224978788
81000,0.9174,0.6893524080298159
```

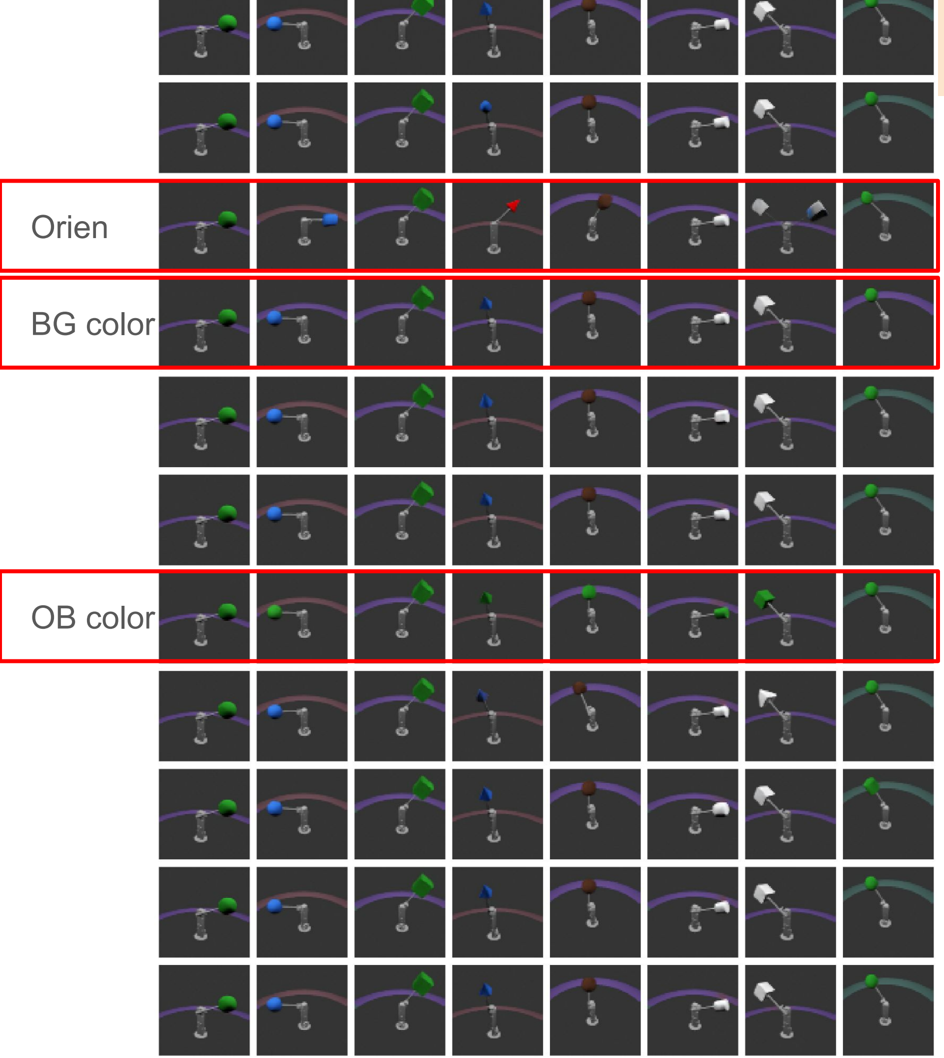
Baseline

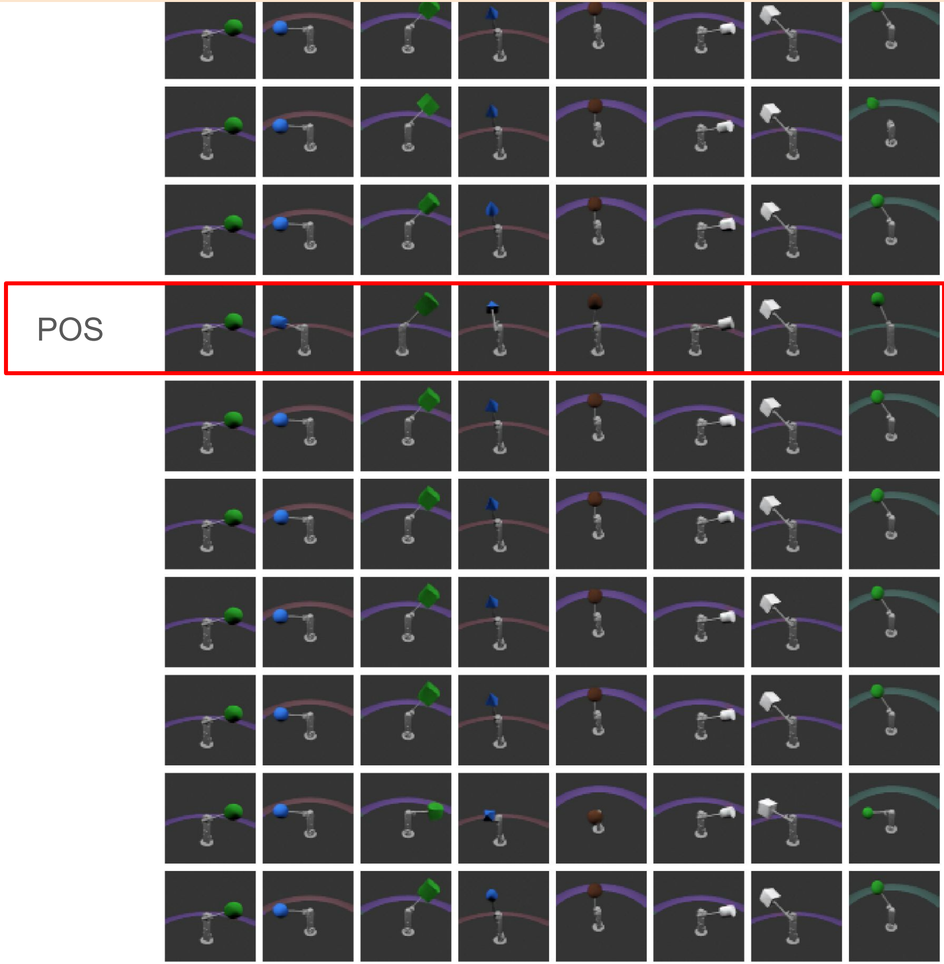


```
timestep,factor_vae_eval_accuracy,dci_dis
8100,0.8108,0.4722781262473835
16200,0.8796,0.6249462197203335
24300,0.9044,0.6605505872422951
32400,0.9204,0.6830328375672906
40500,0.893,0.6284370347603443
48600,0.9126,0.652796622878755
56700,0.9186,0.6707502517943017
64800,0.9262,0.6817782994081243
72900,0.9336,0.7308918357556755
81000,0.9298,0.6793630756441545
```

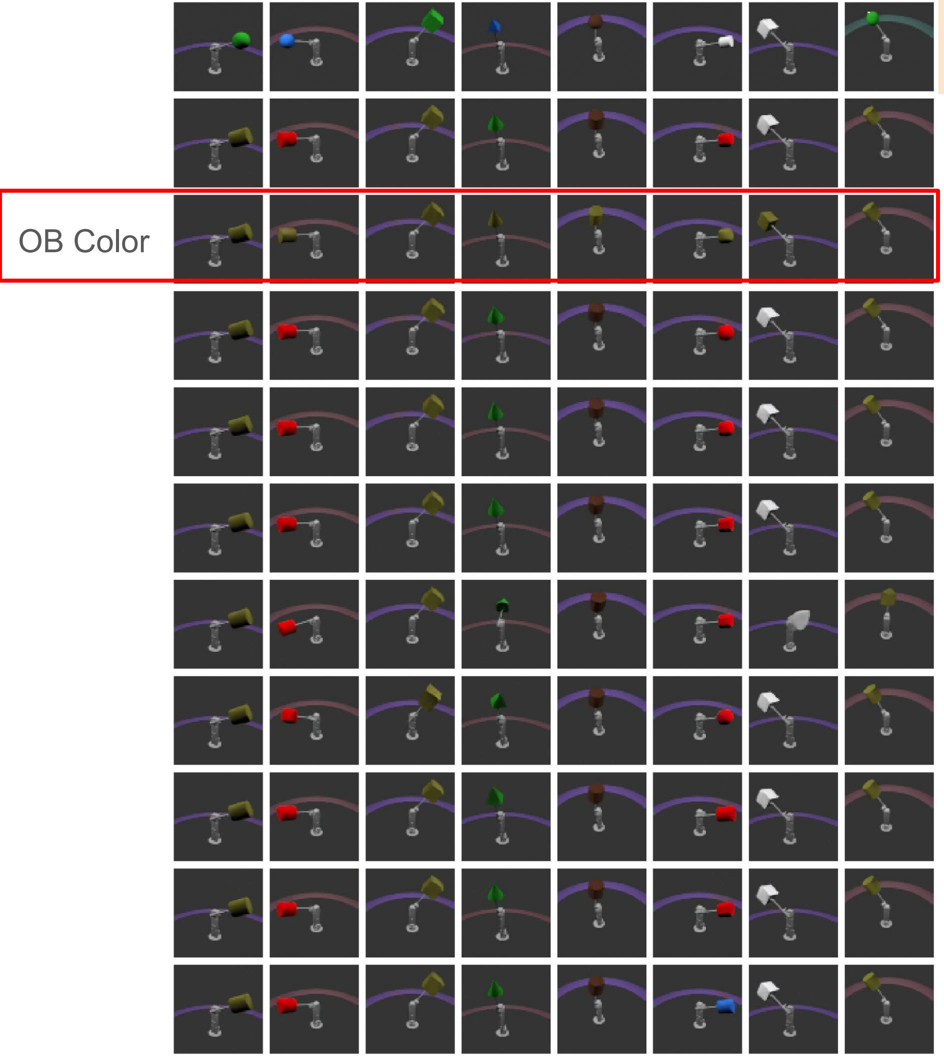
New

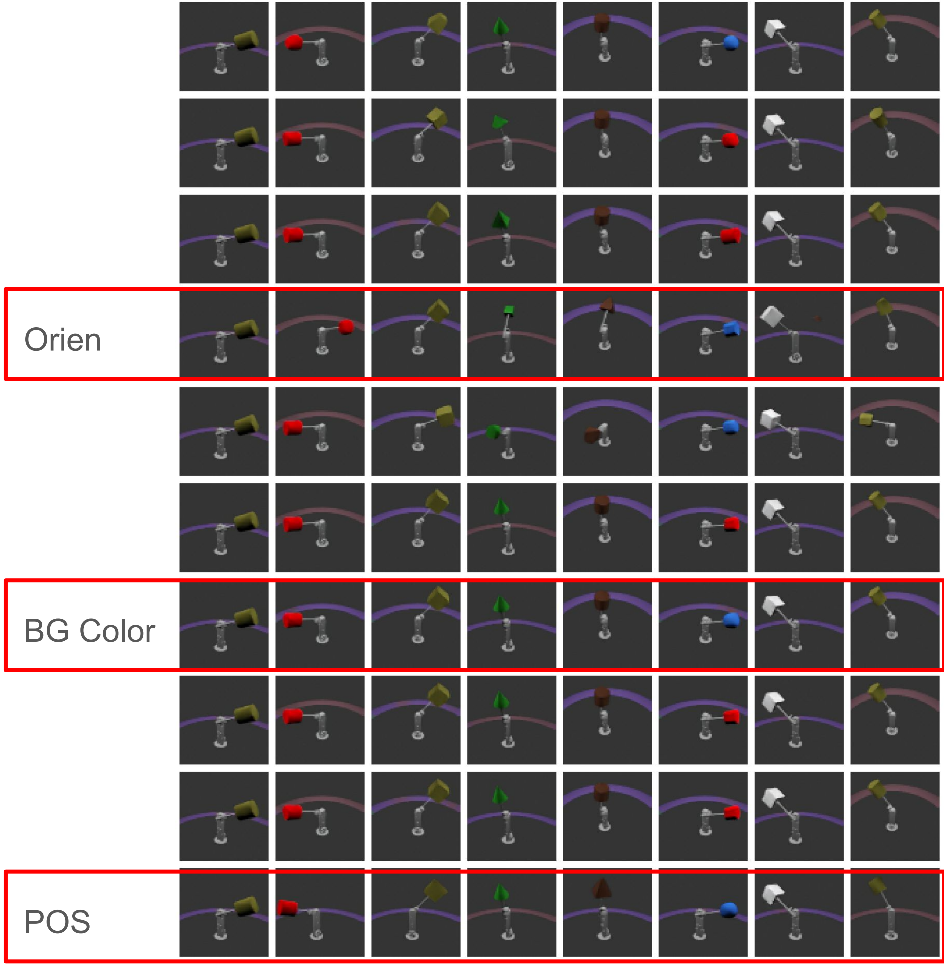
/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-05T06-14-44_mpi3d-vq-4-16-encdiff23/swap_images_last
81000 timesteps, baseline for MPI3D





/mnt/data_7tb/selena/projects/EncDiff/logs/2026-01-10T08-23-34_mpi3d-vq-4-16-encdiff23/swap_images_last
81000 timesteps, w/ concat for MPI3D





Modify Objective Function

Modification could be found by searching “new_loss”, I added that in comments

1. Create ldm/models/diffusion/mcl_utils.py(the same as mcl.py)
2. Modify ldm/models/diffusion/ddpm_enc.py
3. Create configs/mcl