

TEXT 2 IMAGE GENERATION

StackGAN

Maša Cucić
Matematički fakultet
septembar 2024.

TEXT 2 IMAGE

Sadržaj

1. Skup podataka
2. StackGAN
3. CANet
4. Stage1 & Stage 2
5. Trening
6. Test
7. Za buduće
8. Rezultati i zaključak

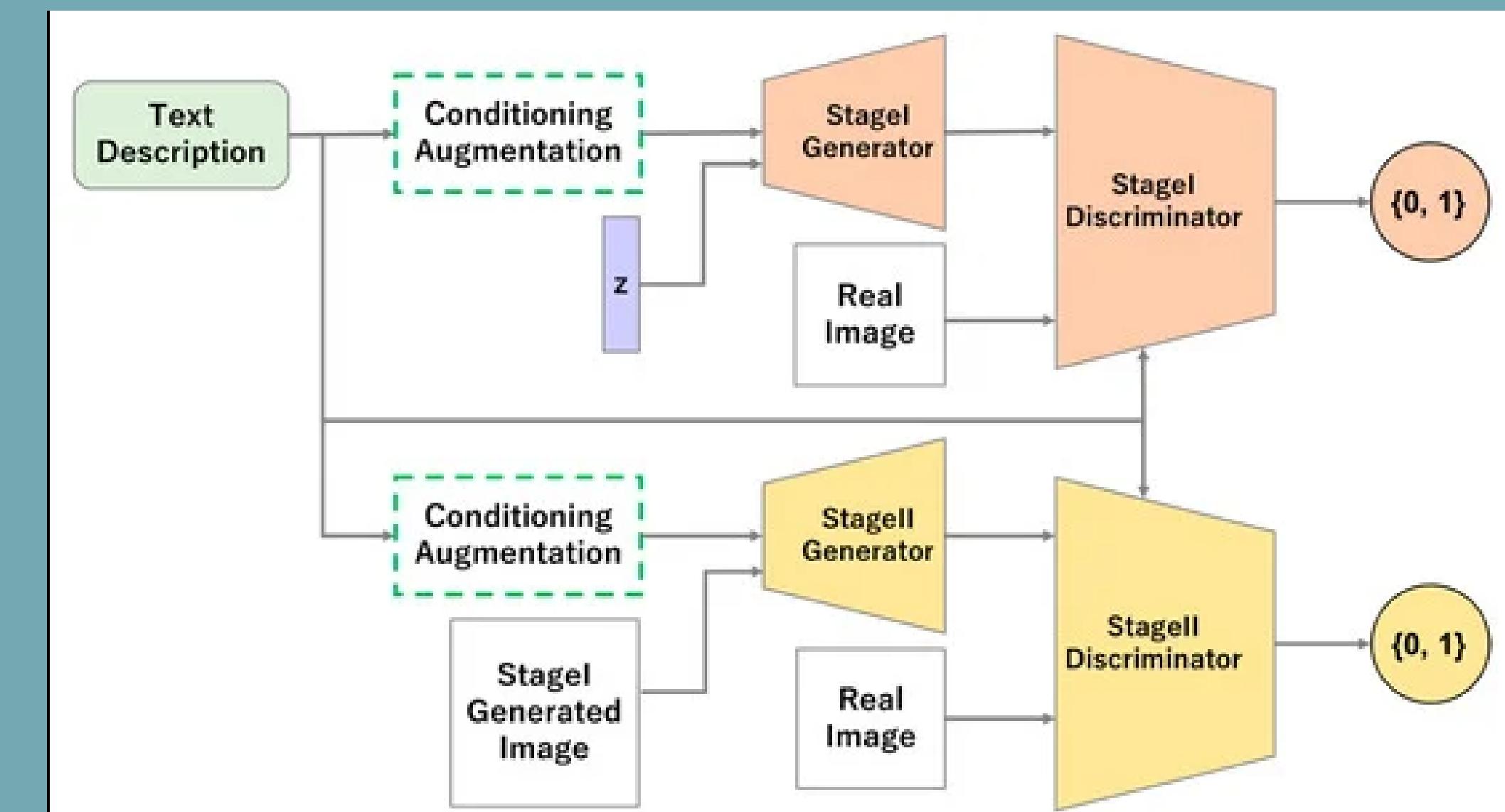
SKUP PODATAKA

- CUB_200_2011 dataset, ptice
- 200 vrsti ptica (40,20,10,1)
- 60 slika po vrsti
- 10 opisa po slici
- train / test 25:75
- char-CNN-RNN-embedding.pickle
- class_info.pickle
- filenames.pickle



STACK GAN

- Generator
- Diskriminator
- Stage 1, Stage 2
- Greške



CA NET

- Conditioning Augmentation Network
- Cilj: enkodiranje tekstualnih embedding-a u latentni prostor koji omogućava stvaranje slika uslovljenih tim tekstualnim opisima
 - mu - srednja vrednost latentne distribucije
 - logvar - log. varijansa latentne distribucije
 - c - konkretan uzorak iz latentnog prostora, dobija se reparamterizacijom

STAGE 1 & STAGE 2

- Generator
- Diskriminator
- CA Net
- 64x64
- 256x256

STAGE 1

Generator

- **text_embedding**
- **noise**
- **CA Net**
- **UpSamplingBlock, nearest neighbor**
- **conv3x3**
- **z_dim, condition_dim, gf_dim**
- **None, fake_img, mu, logvar**

Diskriminatör

- **image**
- **conditions**
- **kodiranje slike**
- **obrada uslovnog vektora**
- **spajanje kodirane slike i obradjenog vektora**
- **finalni output, 0-1**

STAGE 2

Generator

- `text_embedding`
- `kodiranje slike iz prve faze`
- `CA Net`
- `UpSampling`
- `ResidualBlock`
- `conv3x3`
- `z_dim, condition_dim, gf_dim`
- `real_img, fake_img, c, mu, logvar`

Diskriminator

- `image`
- `conditions`
- `kodiranje slike`
- `obrada uslovnog vektora`
- `spajanje kodirane slike i obradjenog vektora`
- `finalni output, 0-1`

TRENING

- 100 epochs
- optimizator Adam, betas=(0.5, 0.999)
- real & fake labels - nizovi jedinica i nula
- noise, z_dim
- Generator Loss - discriminator + kl loss
- Discriminator Loss - pravi, pogrešni i lažni parovi

TRENING

Generator loss

- BCELoss kriterijum
- errD_fake

```
fake_features = netD(fake, cond)
errD_fake = criterion(fake_features.view(-1), real_labels.view(-1))
```

- KL loss - pomaže u približavanju latentne distribucije standardnoj normalnoj distribuciji

```
errG_total = errD_fake + kl_loss*2
```

TRENING

Discriminator loss

- izlaz diskriminatora - verovatnoće
- realni parovi
- pogrešni parovi
- lažni parovi

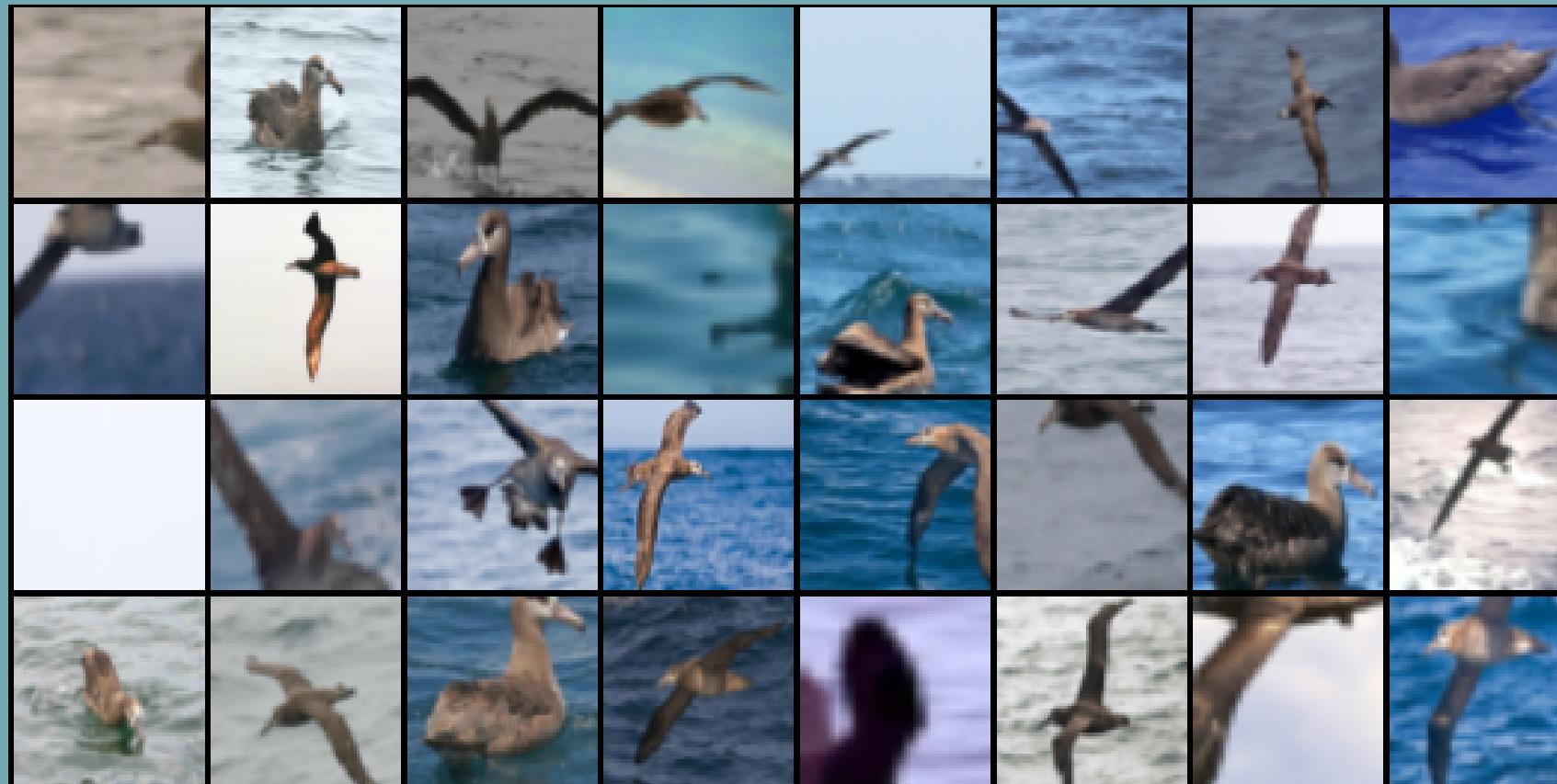
```
errD = errD_real + (errD_fake + errD_wrong)*0.5
```

5

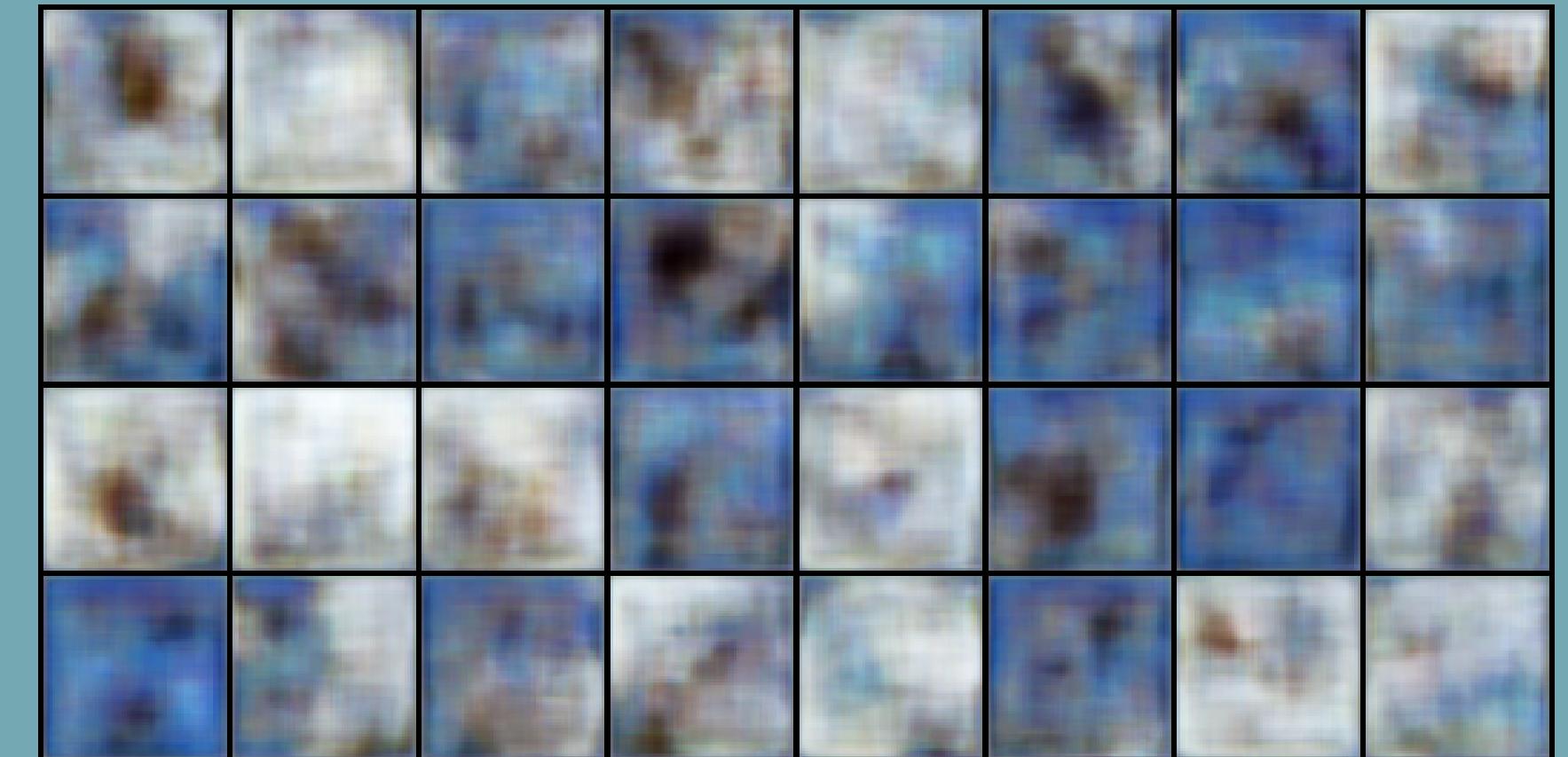
TRENING

1 vrsta

real samples



stage 1

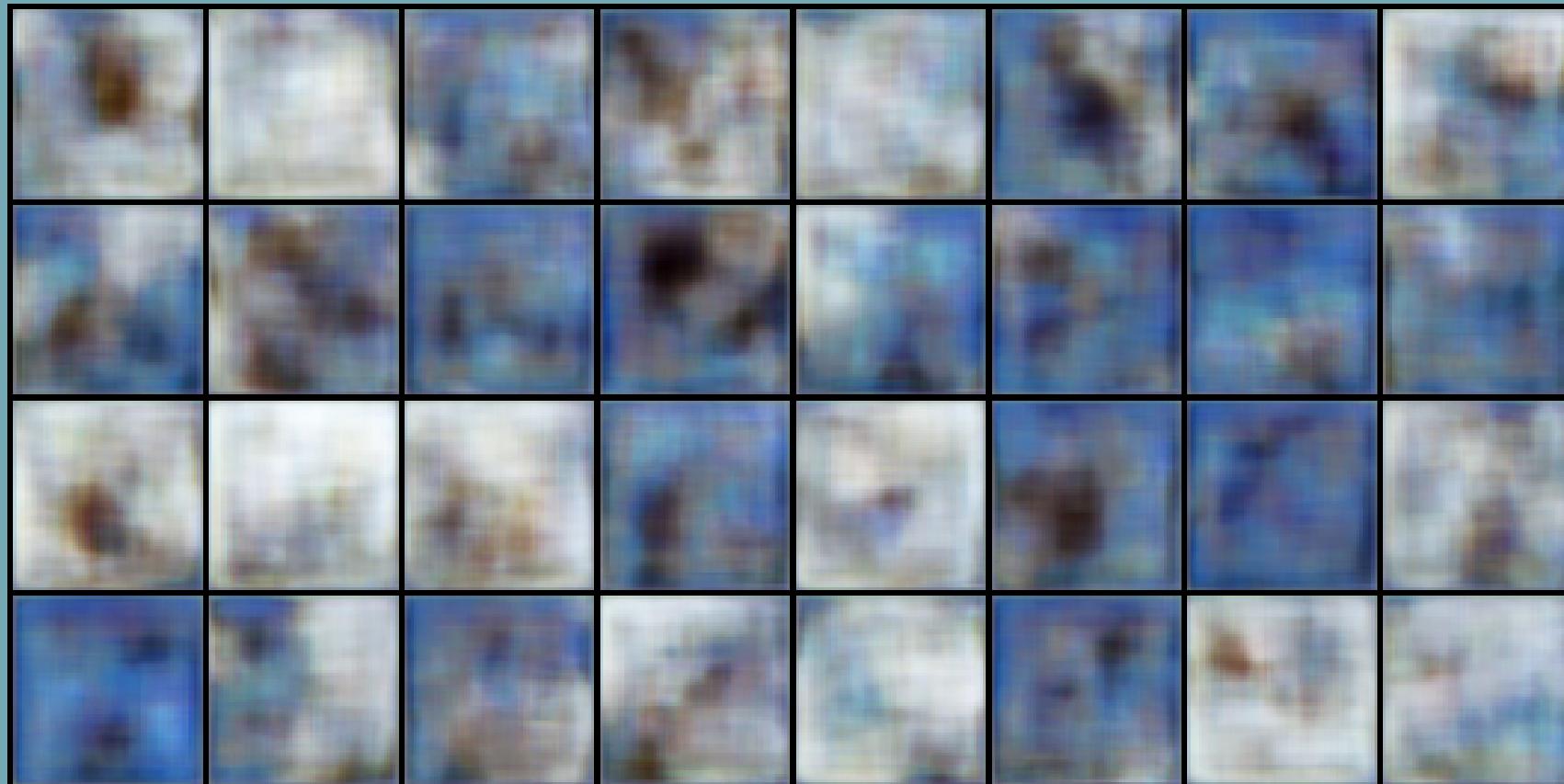


5

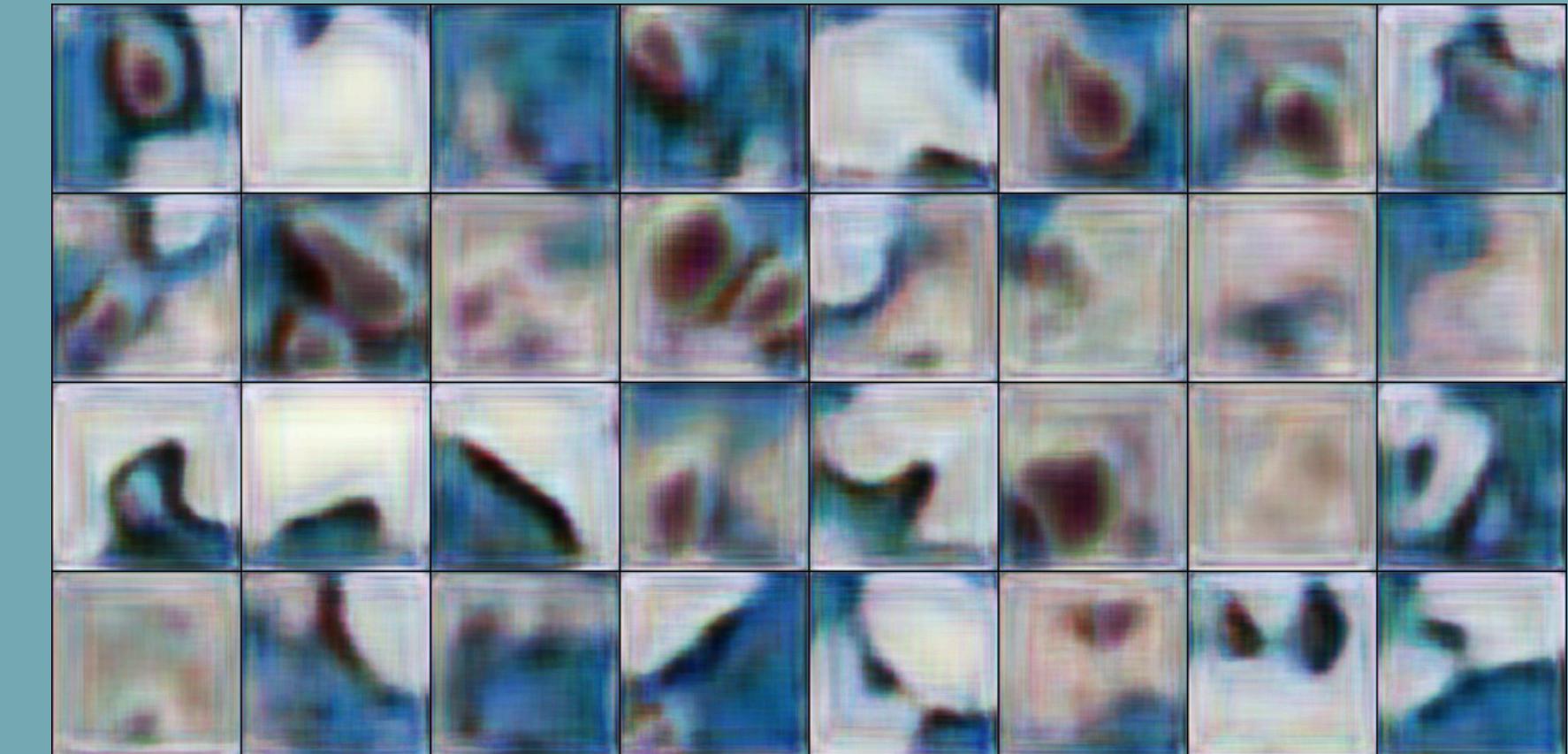
TRENING

1 vrsta

stage 1



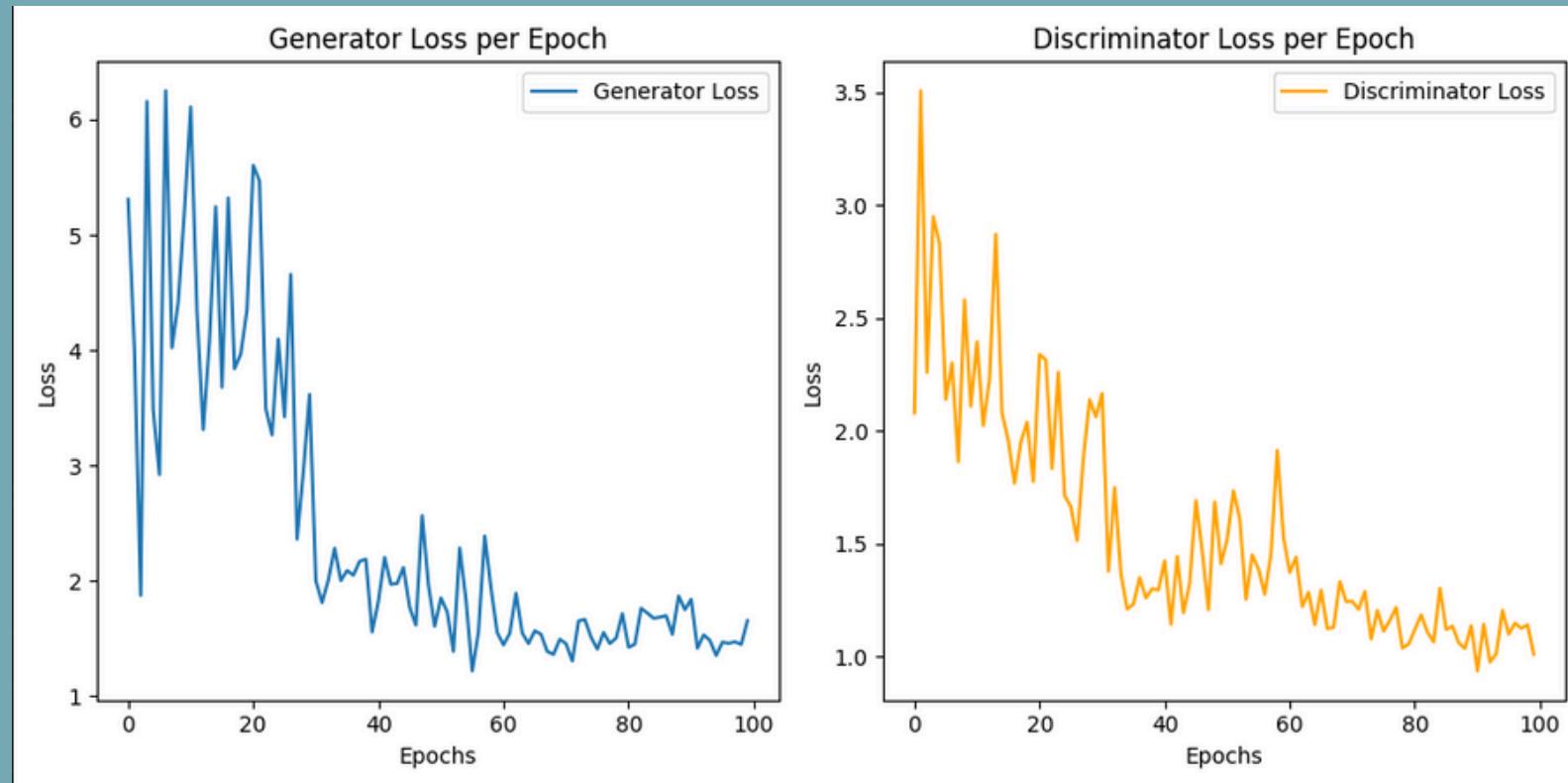
stage 2



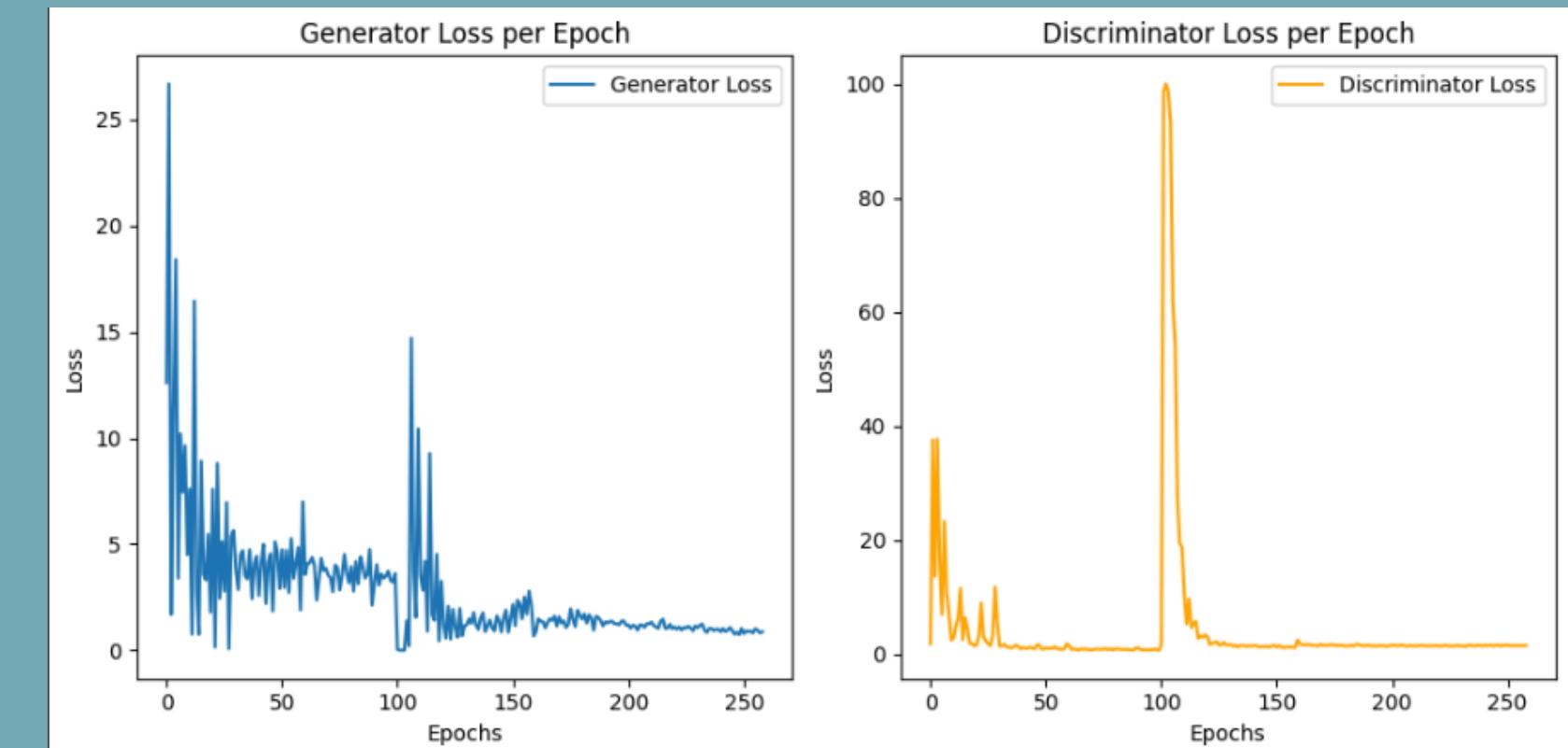
TRENING

1 vrsta

stage 1 greške



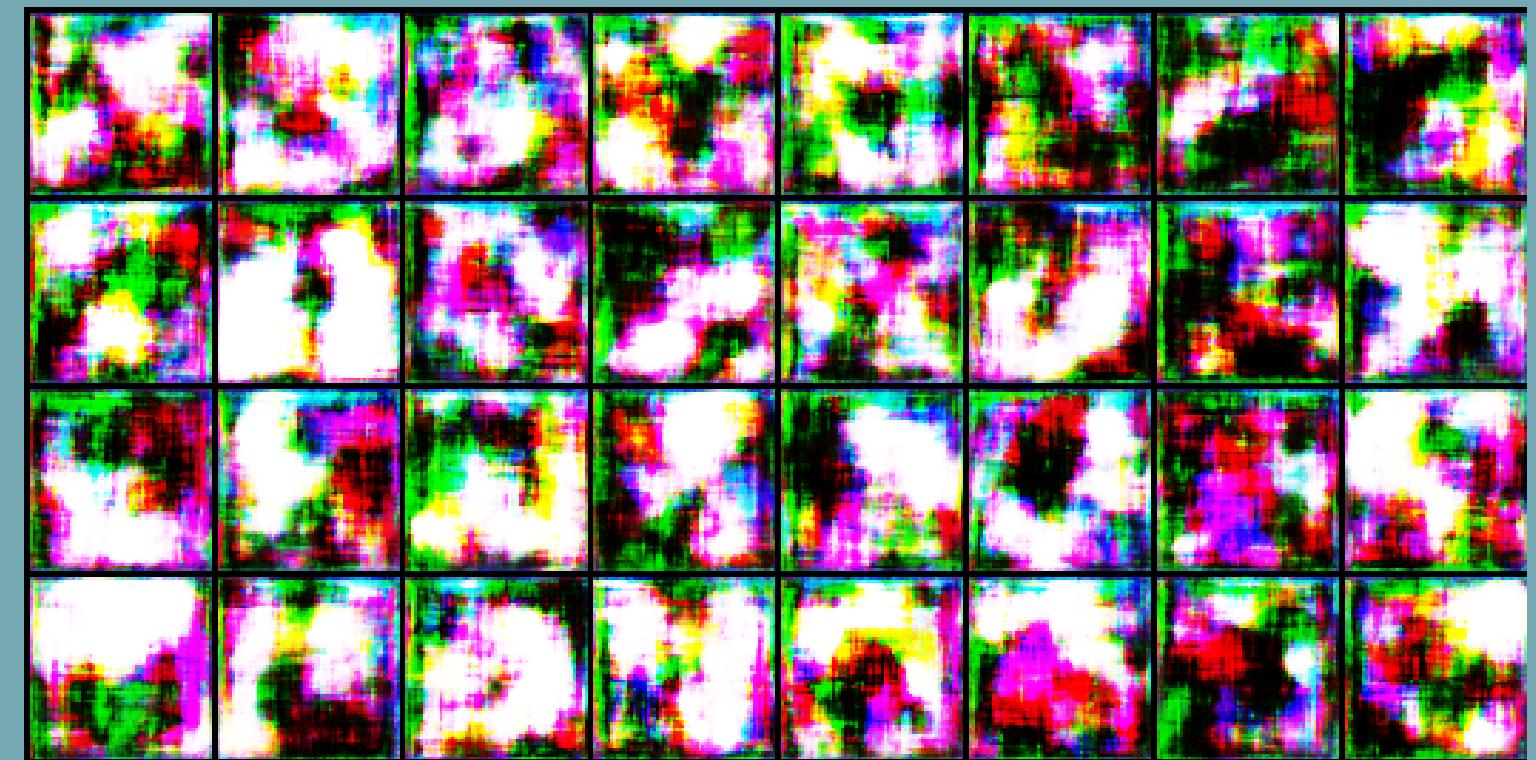
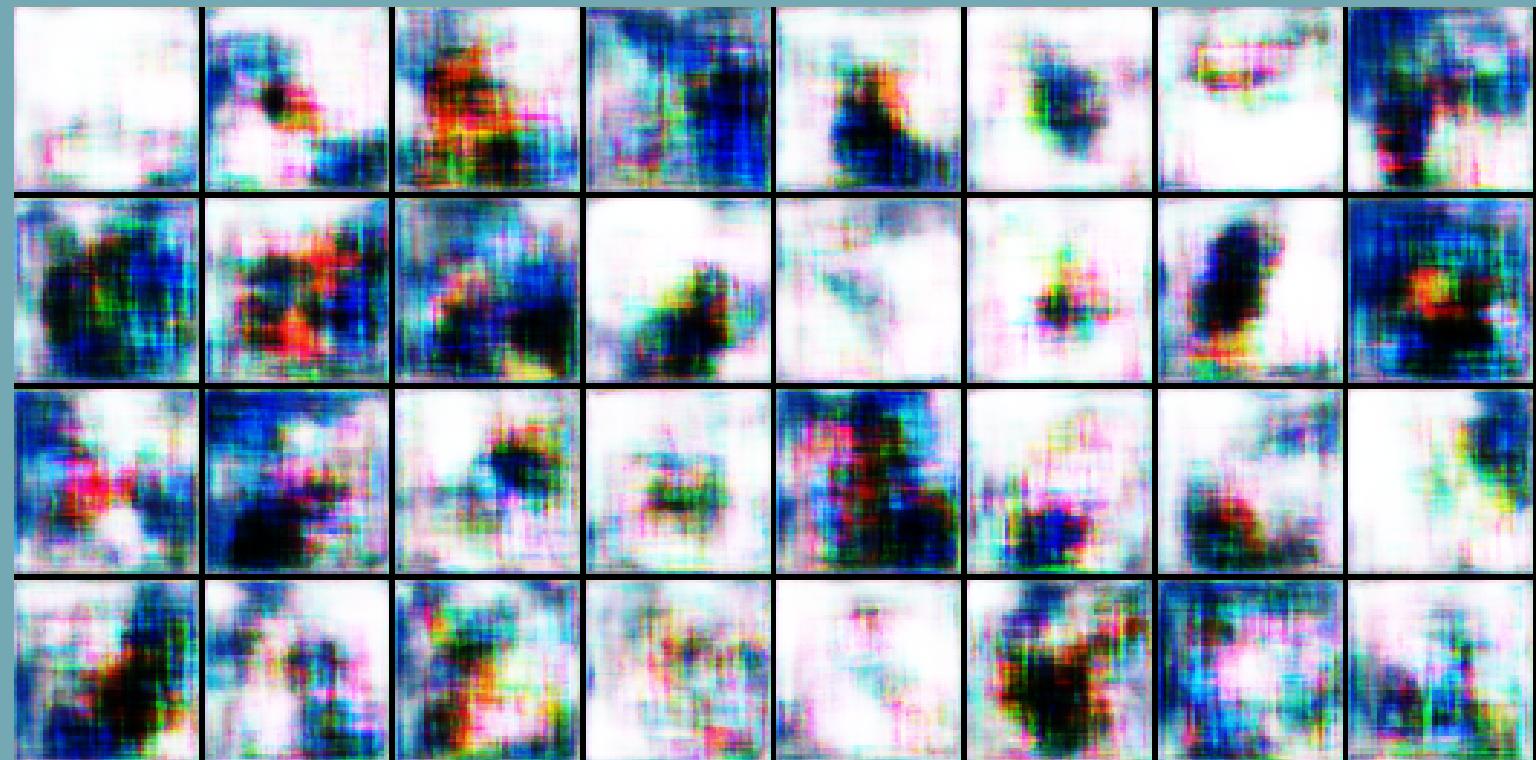
stage 2 greške



TRENING

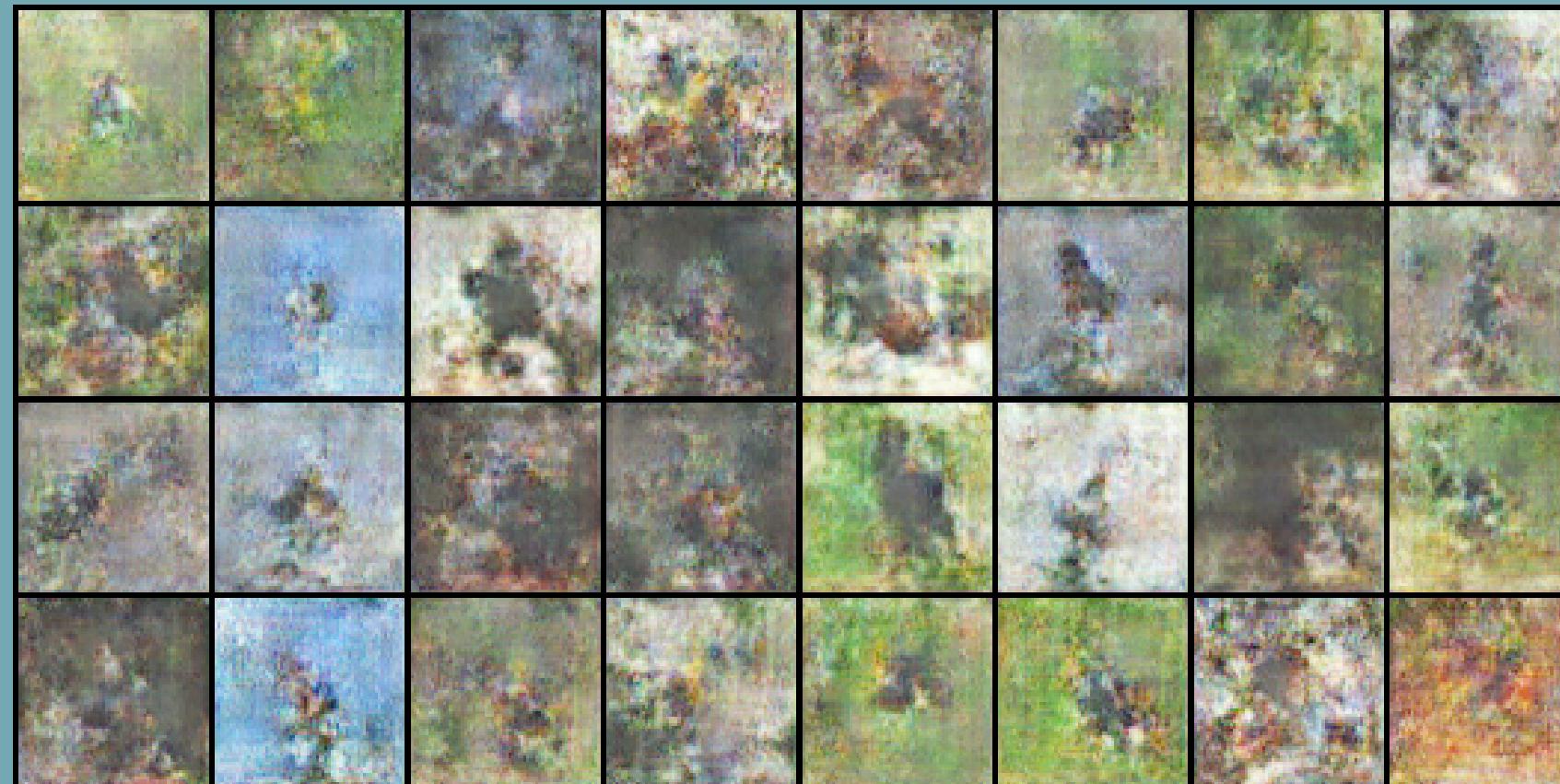
1 vrsta

različite inicijalizacije težina

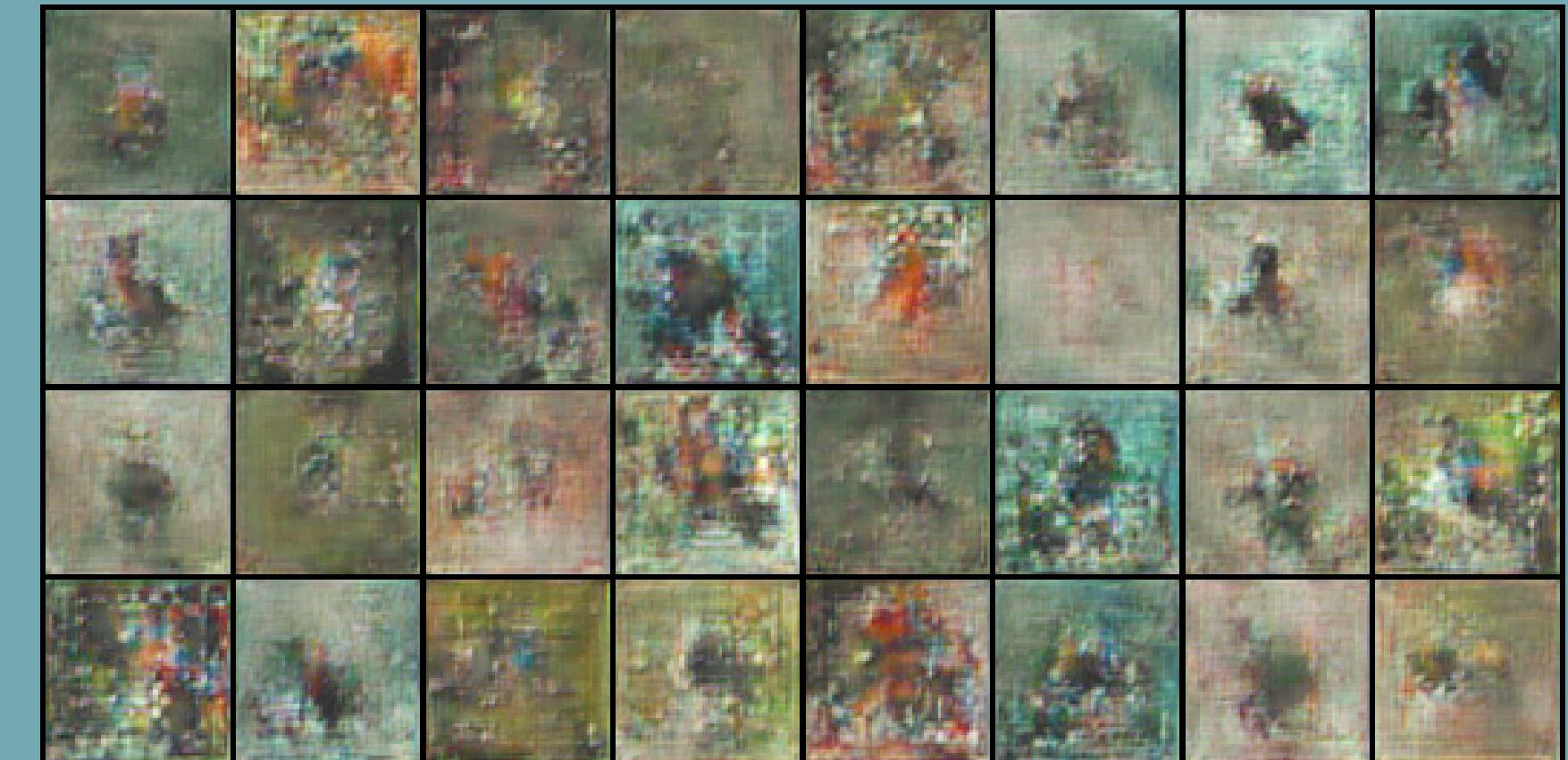


5

TRENING



40 vrsta



20 vrsta

5

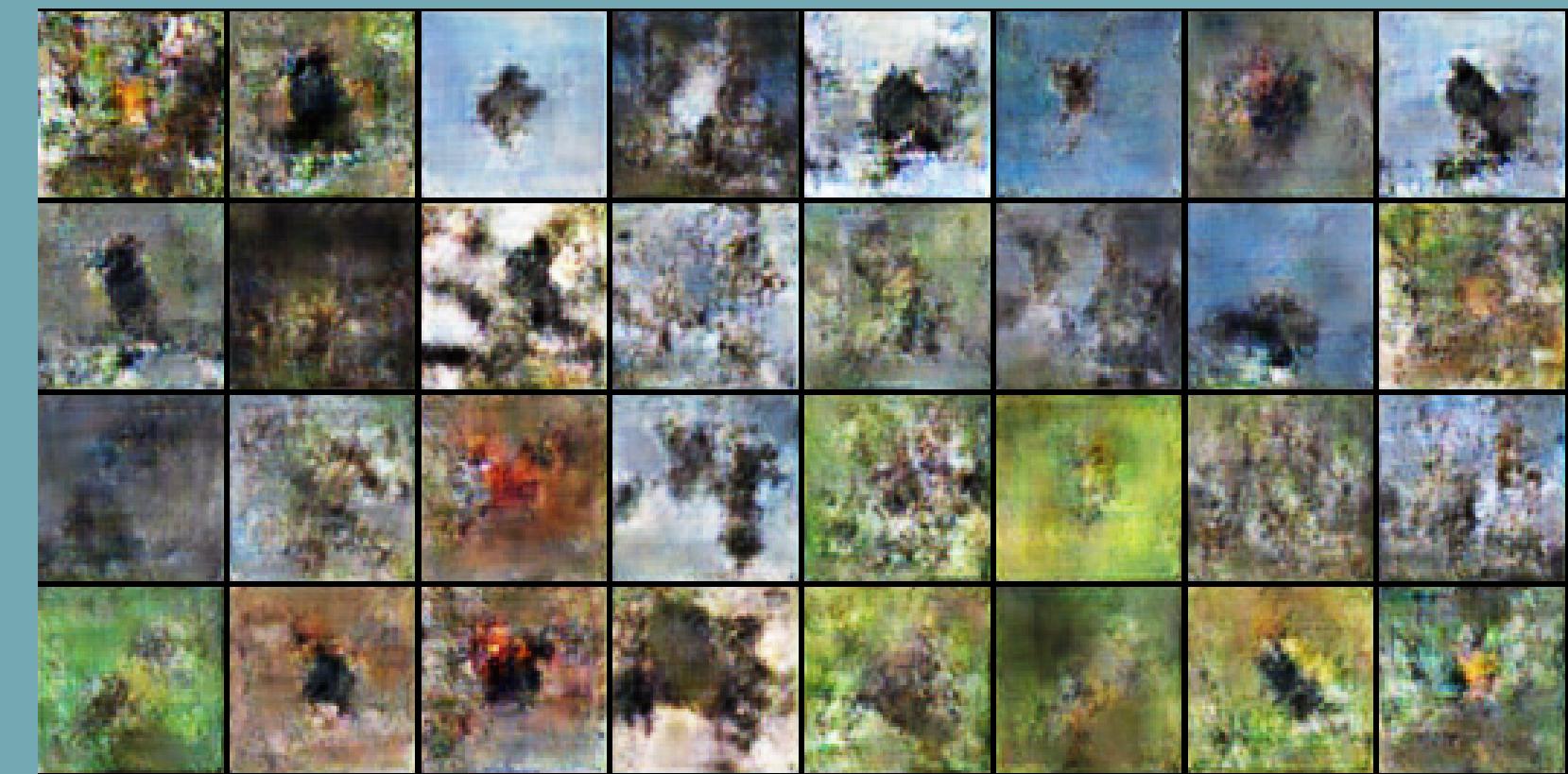
TRENING

40 vrsta

real samples



stage 1



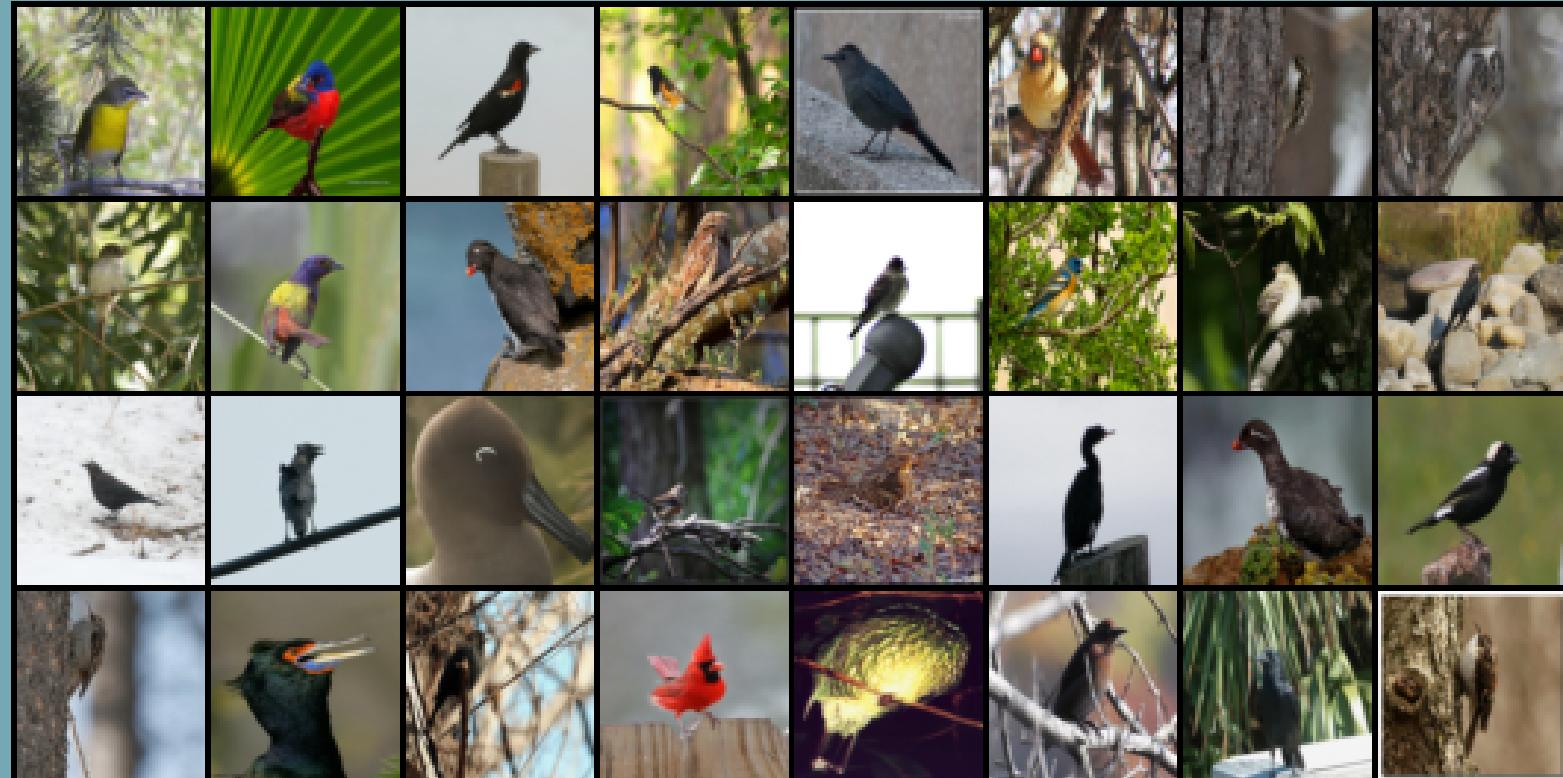
generator lr = 0.0002

5

TRENING

40 vrsta

real samples



stage 1

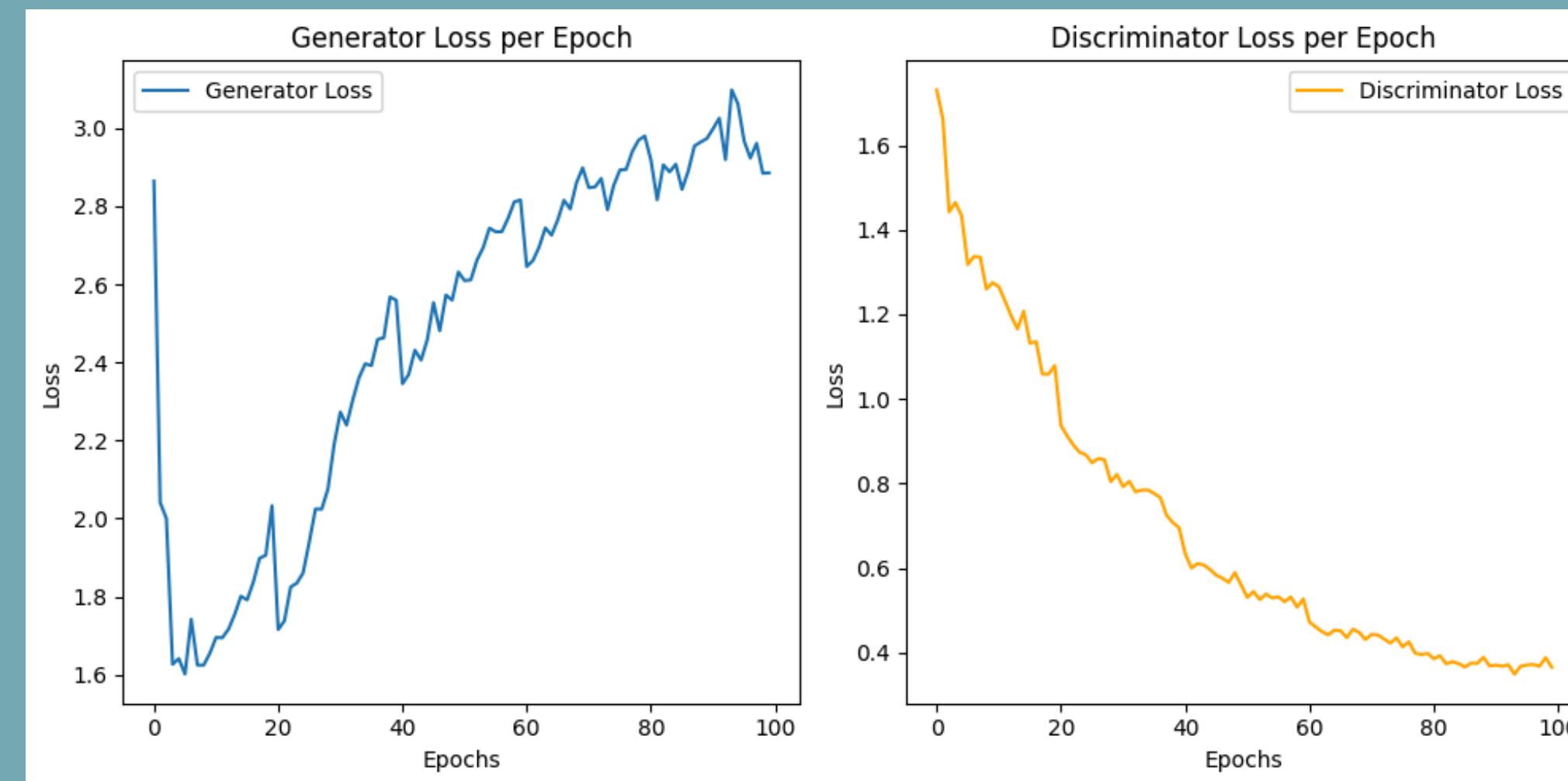


generator lr = 0.0005

TRENING

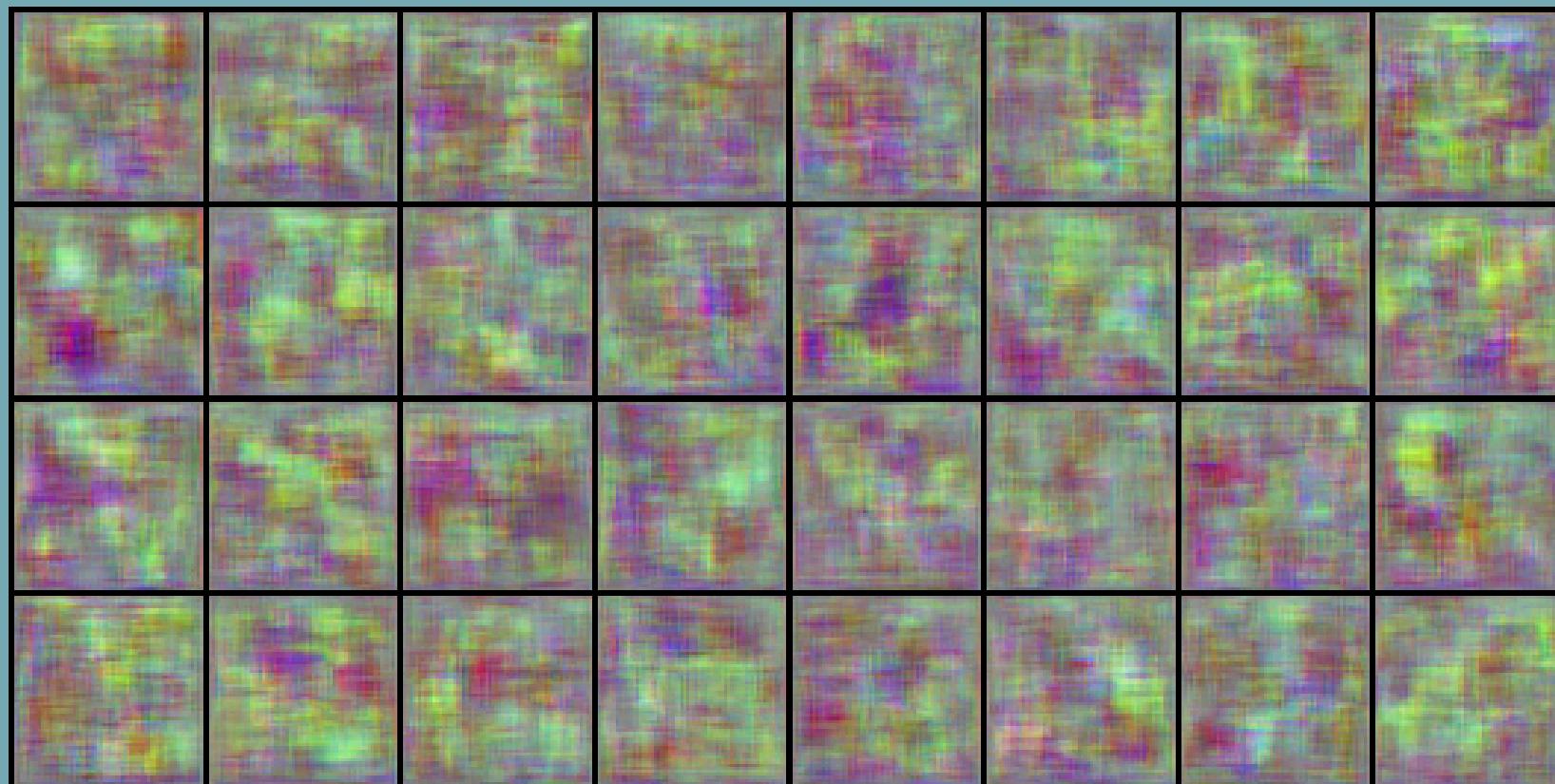
40 vrsta

stage 1 - greške

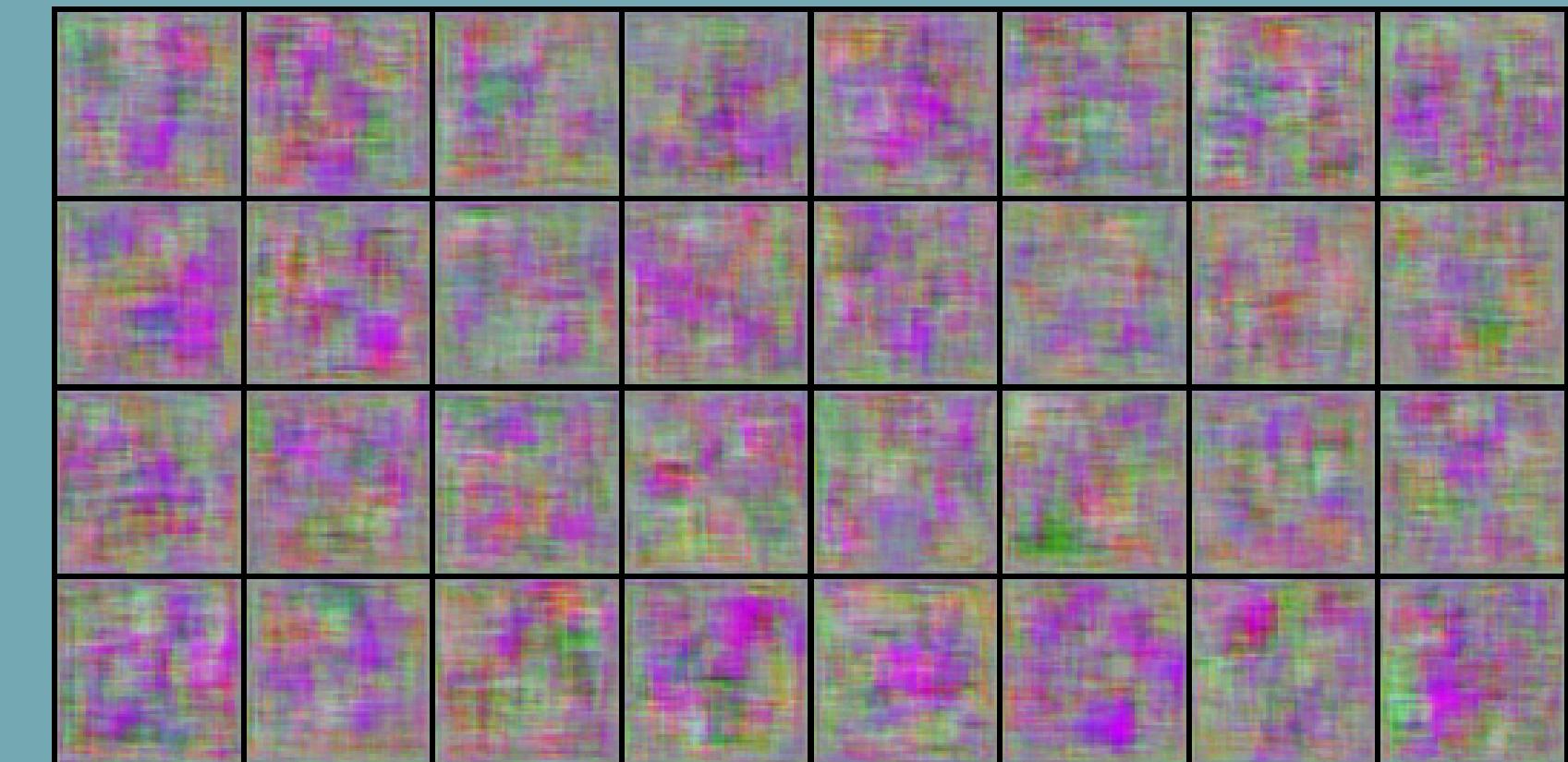


TEST

generisane slike modelom iz prve faze



40 vrsti



20 vrsti

ZA BUDUĆE

- Vizuelna reprezentacija opisima
- Posebni mini batch-evi za prave/lažne/miks slike
- Različite prilagođene inicijalizacije težina generatora i diskriminatora
- Dodatno prilagođavanje arhitektura mreža
- Neuslovljeni diskriminator
- Inception score

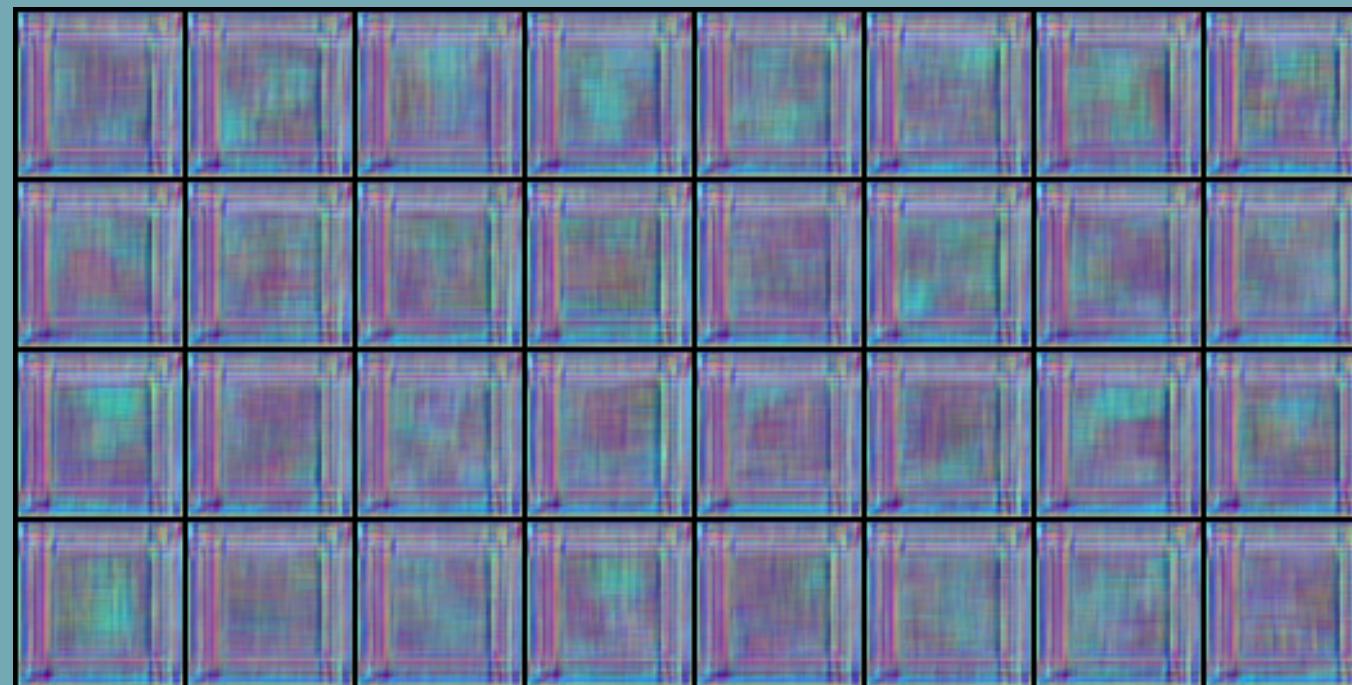
REZULTATI I ZAKLJUČAK

Usklađivanje parametara:

- GF_dim
- DF_dim
- Z_dim
- LR generatora
- LR diskriminatora
- inicijalne težine
- transformacije slika
- optimizatori
- računanje greški

REZULTATI I ZAKLJUČAK

- Mode Collapse
- Balans između generatora i diskriminatora



HVALA NA PAŽNJI

Maša Cucić
Matematički fakultet
septembar 2024.